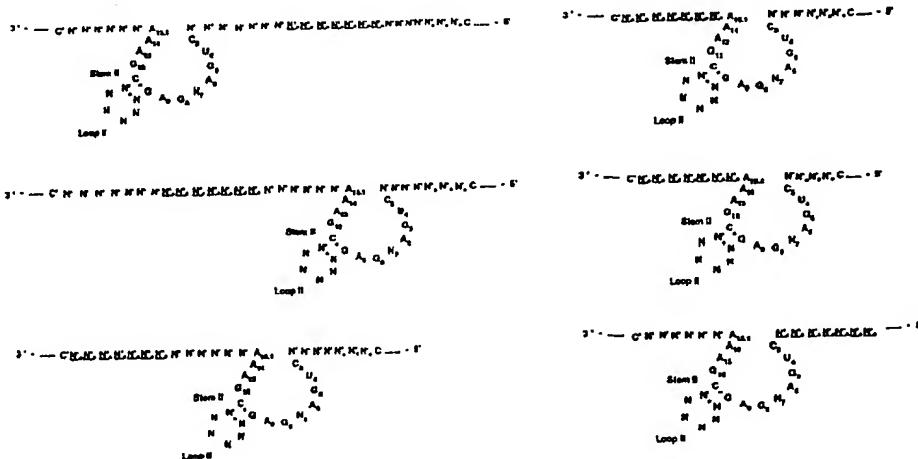




## INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 : <b>C12N 15/11, 9/00, C07H 21/00, A61K 31/70 // C07H 19/00</b>		A2	(11) International Publication Number: <b>WO 99/54459</b>
			(43) International Publication Date: <b>28 October 1999 (28.10.99)</b>
(21) International Application Number: <b>PCT/US99/08547</b>		(74) Agent: <b>WARBURG, Richard, J.; Lyon &amp; Lyon LLP, Suite 4700, 633 West Fifth Street, Los Angeles, CA 90071-2066 (US).</b>	
(22) International Filing Date: <b>19 April 1999 (19.04.99)</b>			
(30) Priority Data: 60/082,404 20 April 1998 (20.04.98) US 09/103,636 23 June 1998 (23.06.98) US		(81) Designated States: AE, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, UZ, VN, YU, ZA, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SL, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).	
(71) Applicant: <b>RIBOZYME PHARMACEUTICALS, INC. [US/US]; 2950 Wilderness Place, Boulder, CO 80301-5411 (US).</b>			
(72) Inventors: <b>THOMPSON, James, D.; 705 Barberry Circle, Lafayette, CO 80026 (US). BEIGELMAN, Leonid; 5530 Colt Drive, Longmont, CO 80503 (US). MCSWIGGEN, James, A.; 4866 Franklin Drive, Boulder, CO 80301 (US). KARPEISKY, Alexander; 420 Vernier Avenue, Lafayette, CO 80026 (US). BELLON, Laurent; 2946 Glenwood Drive, Boulder, CO 80301 (US). REYNOLDS, Mark; 3342 Vermont Place, Pleasanton, CA 94588 (US). ZWICK, Michael; 4138 Joni Lane, Loveland, CO 80207 (US). JARVIS, Thale; 3720 Smuggler Place, Boulder, CO 80303 (US). WOOLF, Tod; Suite 210, 4 Mechanic Street, Natick, MA 01760 (US). HAEBERLI, Peter; 705 7th Street, Berthoud, CO 80513 (US). MATULIC-ADAMIC, Jasenka; 760 South 42nd Street, Boulder, CO 80303 (US).</b>		Published <i>Without international search report and to be republished upon receipt of that report.</i>	
<p align="center"><b>The PTO did not receive the following listed Items(s) <u>from 94-100 pages</u></b></p>			

(54) Title: NUCLEIC ACID MOLECULES WITH NOVEL CHEMICAL COMPOSITIONS CAPABLE OF MODULATING GENE EXPRESSION



## (57) Abstract

The invention features nucleic acid molecules with novel combinations of chemical modifications which are able to modulate gene expression.

**FOR THE PURPOSES OF INFORMATION ONLY**

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece	ML	Mali	TR	Turkey
BG	Bulgaria	HU	Hungary	MN	Mongolia	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MR	Mauritania	UA	Ukraine
BR	Brazil	IL	Israel	MW	Malawi	UG	Uganda
BY	Belarus	IS	Iceland	MX	Mexico	US	United States of America
CA	Canada	IT	Italy	NE	Niger	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NL	Netherlands	VN	Viet Nam
CG	Congo	KE	Kenya	NO	Norway	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NZ	New Zealand	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	PL	Poland		
CM	Cameroon	KR	Republic of Korea	PT	Portugal		
CN	China	KZ	Kazakhstan	RO	Romania		
CU	Cuba	LC	Saint Lucia	RU	Russian Federation		
CZ	Czech Republic	LI	Liechtenstein	SD	Sudan		
DE	Germany	LK	Sri Lanka	SE	Sweden		
DK	Denmark	LR	Liberia	SG	Singapore		
EE	Estonia						

## DESCRIPTION

### Nucleic Acid Molecules With Novel Chemical Compositions Capable Of Modulating Gene Expression

This patent application relates to the patent application entitled, "NUCLEIC ACID 5 MOLECULES WITH NOVEL CHEMICAL COMPOSITIONS CAPABLE OF MODULATING GENE EXPRESSION", U.S.S.N. 60/082,404, which was filed with the U.S. patent and trademark office on April 20, 1998. The earlier patent application listed Thompson *et al.* as inventors.

#### Background Of The Invention

10 This invention relates to novel chemically modified nucleic acid molecules that are capable of modulating gene expression through a variety of mechanisms. Specifically, the invention concerns novel combinations of chemical modifications in an oligonucleotide which enhance nuclease resistance, binding affinity, and/ or potency.

15 The following is a discussion of relevant art, none of which is admitted to be prior art to the present invention.

Since the discovery of the mechanisms underlying gene expression, specifically nucleic acid based transcription and translation, a great deal of effort has been placed on blocking or altering these processes for a variety of purposes, such as understanding biology, gene function, disease processes, and identifying novel therapeutic targets. 20 Approaches involving nucleic acid molecules for modulating gene expression have gained popularity in recent years. For example, nucleic acid molecules have been designed which are capable of binding to specific mRNA sequences by Watson-Crick base pairing interaction and blocking translation (Crooke, 1996, *Medicinal Res. Rev.* 16, 319-344). Another approach involves complexation of DNA with triplex forming oligonucleotides to 25 prevent transcription of bound DNA sequences thereby inhibiting gene expression (Kim *et al.*, 1998, *Biochemistry*. 37, 2299-2304). The interaction of antisense oligonucleotides, 2-5A antisense chimera, or ribozymes with target RNA have been used to prevent gene expression. All of these nucleic acid molecules are highly specific to their matching target sequences and therefore may offer lower toxicity compared to traditional approaches such 30 as chemotherapy.

The use of oligonucleotides for modulation of gene expression generally requires stabilization of oligonucleotides from degradation by nucleases that are present in biological systems. Cellular efficacy may be effected if the nucleic acid molecule is

5 degraded before it reaches its desired target. Chemical modifications of nucleic acid molecules have been found to be advantageous in making them inaccessible to degradation by cellular nucleases. Uhlmann and Peyman, 1990, *Chem. Reviews* 90, 543, review the use of nucleoside modifications to stabilize antisense oligonucleotides. Besides improved stability, chemical modifications have also been shown to increase binding affinity, improve cellular penetration, and enhanced target specificity (Monia *et al.*, 1993, *J. Biol. Chem.* 268, 14514-14522; Wu-Pong, 1994, *BioPharm*, 22-33).

10 One of the most studied and utilized chemical alteration in oligonucleotides has been backbone modifications such as phosphorothioates. Phosphorothioate oligonucleotides are nucleic acid molecules whose phosphodiester linkage has been modified by substituting a sulfur atom in place of an oxygen atom. In addition to increased nucleic acid resistance, phosphorothioate oligonucleotides are substrates for ribonuclease H (RNase H) (Monia, *supra*; Crooke *et al.*, 1995, *Biochem. J.* 3112, 599-608). RNase H is an endonuclease which catalyzes the degradation of RNA in an RNA-DNA heteroduplex 15 (Hostomsky *et al.*, 1993 in *Nucleases*, Linn *et al.*, eds., Cold Spring Harbor Laboratory Press, NY, 341-376). RNA/DNA heteroduplexes, called Okazaki fragments, are formed naturally during DNA replication. Therefore, the normal function of RNase H is to degrade the RNA portion of the heteroduplex to complete DNA replication. In experiments with *E. coli* RNase H, the phosphorothioate oligonucleotide activated the 20 enzyme more efficiently (2-5 fold) compared to a standard phosphodiester containing oligonucleotide (Crooke, 1995, *supra*).

25 Binding of DNA to RNA is not as thermodynamically favorable as an RNA to RNA interaction (Altmann *et al.*, 1996, *Chimia* 50, 168-176). Inoe & Ohtsuka, 1987, *Nucleic Acids Research* 115, 6131, first proposed an oligonucleotide with a central region consisting of oligodeoxynucleotides flanked by 2'-O-methyl modified nucleotide regions. The region of oligodeoxynucleotides in such a chimeric molecule is recognized by RNase H when bound to target RNA; and facilitates cleavage of target RNA by RNase H. (Inoe & Ohtsuka, 1987, *FEBS Lett.* 215, 327; Shibahara & Morisawa, 1987, *Nucleic Acids Res.* 15, 4403). Such chimeric oligonucleotides were proposed to interact with target RNA more 30 stably than an all DNA oligonucleotide.

35 Subsequent developments included the introduction of nuclease resistant modifications of the chimeric oligonucleotides, such as methylphosphonates (Tidd & Gibson, 1988, *Anticancer Drug Design* 3, 117), phosphorothioates (Agrawal & Pederson, 1990, *Proc Nat. Acad. Sci. USA* 87, 1407), and phosphoramidates (Potts & Runyun, 1991, *Proc Nat. Acad. Sci. USA* 88, 1516). Additionally, the flanking sequences have been

modified with 2'-*O*-methyl and 2'-*F*-modifications (Cook, 1993, *Antisense Research and Applications*, CRC Press, 150-181).

Agrawal et al., US Patent No. 5,652,355, describe a phosphorothioate-containing nucleic acid molecule with at least two 2'-*O*-methyl modifications on the 5' and 3' ends.

5 Agrawal, US Patent No. 5,652,356, describes an oligonucleotide which consists of a region of 2'-*O*-substituted oligonucleotide located between two oligodeoxyribonucleotide regions. The DNA regions of this nucleic acid molecule consists of phosphorothioate modifications at every position.

10 Cook et al., US Patent No. 5,623,065, describe the use of a nucleic acid molecule which contains an RNase H cleavable region flanked by certain specifically modified nucleotides, for inhibition of gene expression of a ras gene.

Cook et al., US Patent No. 5,587,362, describe a nucleic acid molecule having "substantially chirally pure inter-sugar linkages", for modulation of gene expression.

15 Ohtsuka et al., US Patent No. 5,013,830, describe mixed oligomers having a DNA region and a 2'-*O*-methyl modified region, useful for modulation of gene expression.

Walder et al., US Patent No. 5,491,133, describe a method for modulating gene expression using chimeric oligonucleotides with 3'-phosphodiester linkage modifications.

20 Cohen et al., US Patent No. 5,276,019, and Cohen et al., US Patent No. 5,264,423 describe the use of oligodeoxynucleotides of no more than 32 nucleotides in length, containing at least one phosphorothioate internucleoside linkage which are capable of preventing foreign nucleic acid replication.

Cohen et al., US Patent No. 5,286,717, describe an oligodeoxyribonucleotide with at least one phosphorothioate modification capable of inhibiting oncogenes.

25 Sproat et al., US Patent No. 5,334,711, describe 2'-*O*-R modified hammerhead and hairpin ribozymes, where R IS ALKYL, ALKYNYL OR ALKENYL.

Crooke et al., 1996, *Exp. Opin. Ther. Patents* 6, 855, list and discuss various patents and PCT publications in the field of antisense technology.

Sproat et al., US Patent No. 5,678,731, describe 2'-*O*-R modified oligonucleotides where R IS ALKYL, ALKYNYL OR ALKENYL.

30 Usman et al., US Patent No. 5,652,094, describe enzymatic nucleic acid molecules which include nucleic acid analogues or deoxyribonucleotides.

Joyce, International Publication No. WO 96/17086, describes a DNA enzyme capable of cleaving RNA.

35 Rossi et al., US Patent No. 5,144,019, describe chimeric hammerhead ribozymes with the binding arms and stem II region modified with deoxyribonucleotides.

Molecules have also been devised which include non-nucleotides capable of binding to nucleic acid. These peptide nucleic acid (PNA) molecules bind by Watson-Crick base-pairing and may also function through an antisense mechanism. These molecules have been used to augment hammerhead ribozyme activity by altering the 5 structure of target RNAs and increasing accessibility of cleavage sites (Jankowsky *et al.*, 1997, *Nucleic Acids Research* 25, 2690-2693).

Summary of the Invention

This invention relates to novel nucleic acid molecules which are useful for modulation of gene expression. The nucleic acid molecule of the instant invention are 10 distinct from other nucleic acid molecules known in the art. Specifically, the nucleic acid molecules of the present invention have novel combinations of chemical modifications and are capable of binding to RNA or DNA to facilitate modulation of gene expression. These novel combinations of chemical modifications may be used to form antisense oligonucleotides, triplex forming oligonucleotides, 2-5A antisense chimera, and enzymatic 15 nucleic acid molecules.

In a preferred embodiment, the invention features a nucleic acid molecule having the following formulae:

Formula I:



20 Formula II:



Formula III:

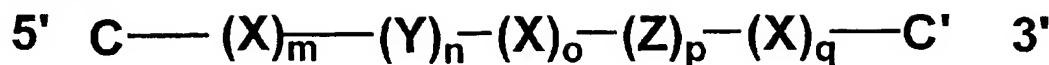


In a preferred embodiment, the invention features an enzymatic nucleic acid molecule having the formula:

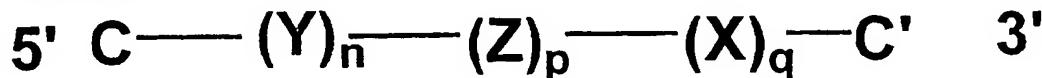
Formula IV:



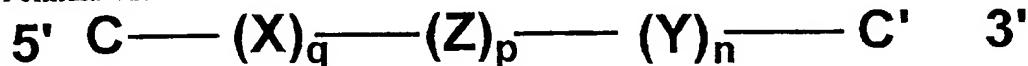
Formula V:



Formula VI:



5 Formula VII



In each of the above formula (I-VII), X represents independently a nucleotide which may be same or different; where m and o are integers independently greater than or equal to 4 and preferably less than about 100, more specifically 5, 6, 7, 8, 9, 10, 11, 12, 15, 10 or 20; r is an integer greater than or equal to four, more specifically 5, 6, 7, 10, 15, or 20; the nucleic acid molecule may be symmetric (m = O) or asymmetric (m ≠ O); (X)<sub>m</sub>, (X)<sub>o</sub>, and (X)<sub>q</sub> are oligonucleotides which are of sufficient length to stably interact independently with a target nucleic acid molecule (the target can be an RNA, DNA or RNA/DNA mixed polymers); Y represents independently a deoxyribonucleotide which 15 may be same or different; n is an integer greater than or equal to 4, specifically 5, 6 7, 8, 9, 10, 11, or 12; Z represents an oligonucleotide including nucleotides capable of facilitating the cleavage of a target sequence; p is of length greater than or equal to 4 but less than 100, preferably 5, between 10-20, specifically 25-55, specifically between 30-45, more 20 specifically 35-50; q is an integer greater than or equal to 0, preferably 1, 2, 3, 4, 5, 6, 7, 8, 10, 15, 20; — represents a chemical linkage (e.g. a phosphate ester linkage, amide linkage or others known in the art); and each (X)<sub>m</sub>, (X)<sub>o</sub>, (X)<sub>r</sub>, (X)<sub>q</sub>, and/or (Y)<sub>n</sub> independently comprise phosphorothioate linkages, more specifically each (X)<sub>m</sub>, (X)<sub>o</sub>, (X)<sub>r</sub>, (X)<sub>q</sub>, and/or (Y)<sub>n</sub> independently comprise at least one phosphodiester linkage and one phosphorothioate linkage; each C and C' independently represents a cap structure which 25 may independently be present or absent; and (Z)<sub>p</sub> may optionally include a phosphorothioate linkage. The nucleotides in the each of the formula I-VII are unmodified or modified at the sugar, base, and/or phosphate as known in the art.

30 Preferably, each of X represents independently a nucleotide which may be same or different; where m and o are integers independently greater than or equal to 5; (X)<sub>m</sub> and (X)<sub>o</sub> are oligonucleotides which are of sufficient length to stably interact independently with a target nucleic acid molecule; each (X)<sub>r</sub> comprises independently at least one phosphodiester linkage and one phosphorothioate linkage; Y represents independently a

deoxyribonucleotide which may be same or different; (Y)<sub>n</sub> is an oligonucleotide which is of sufficient length to stably interact independently with a target nucleic acid molecule; n is an integer greater than or equal to 4; each (X)<sub>m</sub>, and (X)<sub>o</sub> comprise independently at least one phosphodiester linkage and one phosphorothioate linkage; (Y)<sub>n</sub> comprises a 5 phosphorothioate linkage or a phosphorodithioate linkage or a 5'-S-phosphorothioate, or 5'-S-phosphorodithioate, or a 3'-S-phosphorothioate or a 3'-S-phosphorodithioate linkage or a mixture thereof; and each C and C' independently represents a cap structure which may independently be present or absent.

By "nucleotide" as used herein is as recognized in the art to include natural bases (standard), and modified bases well known in the art. Such bases are generally located at the 1' position of a sugar moiety. Nucleotide generally comprise a base, sugar and a phosphate group. The nucleotides can be unmodified or modified at the sugar, phosphate and/or base moiety, (also referred to interchangeably as nucleotide analogs, modified nucleotides, non-natural nucleotides, non-standard nucleotides and other; see for example, 10 Usman and McSwiggen, *supra*; Eckstein *et al.*, International PCT Publication No. WO 92/07065; Usman *et al.*, International PCT Publication No. WO 93/15187; Uhlman & Peyman, *supra*) all are hereby incorporated by reference herein). There are several 15 examples of modified nucleic acid bases known in the art and has recently been summarized by Limbach *et al.*, 1994, *Nucleic Acids Res.* 22, 2183. Some of the non-limiting examples of base modifications that can be introduced into nucleic acids include, 20 inosine, purine, pyridin-4-one, pyridin-2-one, phenyl, pseudouracil, 2, 4, 6-trimethoxy benzene, 3-methyl uracil, dihydrouridine, naphthyl, aminophenyl, 5-alkylcytidines (e.g., 25 5-methylcytidine), 5-alkyluridines (e.g., ribothymidine), 5-halouridine (e.g., 5-bromouridine) or 6-azapyrimidines or 6-alkylpyrimidines (e.g. 6-methyluridine), 30 propyne, and others (Burgin *et al.*, 1996, *Biochemistry*, 35, 14090; Uhlman & Peyman, *supra*). By "modified bases" in this aspect is meant nucleotide bases other than adenine, guanine, cytosine and uracil at 1' position or their equivalents; such bases may be used at any position, for example, within the catalytic core of an enzymatic nucleic acid molecule and/or in the substrate-binding regions of the nucleic acid molecule.

30 By "unmodified nucleotide" is meant a nucleotide with one of the bases adenine, cytosine, guanine, thymine, uracil joined to the 1' carbon of  $\beta$ -D-ribo-furanose.

By "modified nucleotide" is meant a nucleotide which contains a modification in the chemical structure of an unmodified nucleotide base, sugar and/or phosphate.

35 By "sufficient length" is meant an oligonucleotide of greater than or equal to 4 nucleotides.

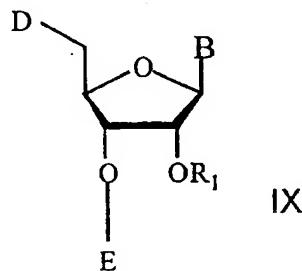
By "stably interact" is meant, interaction of the oligonucleotides with target nucleic acid (e.g., by forming hydrogen bonds with complementary nucleotides in the target under physiological conditions). The interaction is stable either alone or in conjunction with (Y)<sub>n</sub> and (Z)<sub>p</sub> where applicable.

5 By "chimeric nucleic acid molecule" or "chimeric oligonucleotide" is meant that, the molecule may be comprised of both modified or unmodified DNA or RNA.

10 By "cap structure" is meant chemical modifications which have been incorporated at the terminus of the oligonucleotide. These terminal modifications protect the nucleic acid molecule from exonuclease degradation, and may help in delivery and/or localization 10 within a cell.

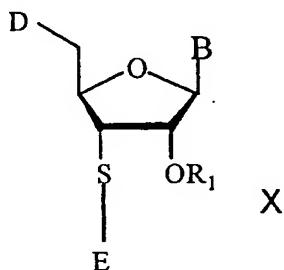
15 In another preferred embodiment (X)<sub>m</sub>, (X)<sub>o</sub>, (X)<sub>q</sub>, (Y)<sub>n</sub> and/or (Z)<sub>p</sub> independently include modifications selected from a group comprising 2'-O-alkyl (e.g. 2'-O-allyl; Sproat *et al.*, *supra*); 2'-O-alkylthioalkyl (e.g. 2'-O-methylthiomethyl; Karpeisky *et al.*, 1998, *Nucleosides & Nucleotides* 16, 955-958); L-nucleotides (Tazawa *et al.*, 1970, *Biochemistry* 3499; Ashley, 1992, *J. Am. Chem. Soc.* 114, 9731; Klubmann *et al.*, 1996, *Nature Biotech* 14, 1112); 2'-C-alkyl (Beigelman *et al.*, 1995, *J. Biol. Chem.* 270, 25702); 1-5-Anhydrohexitol; 2,6-diaminopurine (Strobel *et al.*, 1994, *Biochem.* 33, 13824-13835); 2'-(N-alanyl) amino-2'-deoxynucleotide; 2'-(N-beta-alanyl) amino; 2'-deoxy-2'-(lysyl) amino; 2'-O-amino (Karpeisky *et al.*, 1995, *Tetrahedron Lett.* 39, 1131); 2'-deoxy-2'-(N-histidyl) amino; 5-methyl (Strobel, *supra*); 2'-(N-beta-carboxamidine-beta-alanyl) amino; 2'-deoxy-2'-(N-beta-alanyl) (Matulic-Adamic *et al.*, 1995, *Bioorg. & Med. Chem. Lett.* 5,2721-2724); xylofuranosyl (Rosemeyer *et al.*, 1991, *Helvetica Chem. Acta*, 74, 748; Seela *et al.*, 1994, *Helvetica Chem. Acta*, 77, 883; Seela *et al.*, 1996, *Helvetica Chem. Acta*, 79, 1451).

20 25 In a preferred embodiment, the invention features a nucleic acid molecule of any of formula I-VII, where each X and/or Z, independently include a nucleotide modification having formula IX:



Where, each B is independently a modified or an unmodified nucleic acid base; R1 is independently a fluoroalkyl or an alkylthiofluoroalkyl; E is independently a phosphorus-containing group; and D is independently an O, blocking group or a phosphorus-containing group.

5 In another preferred embodiment, the invention features a nucleic acid molecule of any of formula I-VII, where each X and/or Z, independently include a nucleotide modification having formula X:



10 Wherein, each B is independently a modified or an unmodified nucleic acid base; R1 is independently an alkyl, an alkylthioalkyl, a fluoroalkyl or an alkylthiofluoroalkyl; E is independently a phosphorus-containing group; and D is independently an O, blocking group or a phosphorus-containing group.

An "alkyl" group refers to a saturated aliphatic hydrocarbon, including straight-chain, branched-chain, and cyclic alkyl groups. Preferably, the alkyl group has 1 to 12 carbons. More preferably it is a lower alkyl of from 1 to 7 carbons, more preferably 1 to 4 carbons. The alkyl group may be substituted or unsubstituted. When substituted the substituted group(s) is preferably, hydroxyl, cyano, alkoxy, =O, =S, NO<sub>2</sub> or N(CH<sub>3</sub>)<sub>2</sub>, amino, or SH. The term also includes alkenyl groups which are unsaturated hydrocarbon groups containing at least one carbon-carbon double bond, including straight-chain, branched-chain, and cyclic groups. Preferably, the alkenyl group has 1 to 12 carbons. More preferably it is a lower alkenyl of from 1 to 7 carbons, more preferably 1 to 4 carbons. The alkenyl group may be substituted or unsubstituted. When substituted the substituted group(s) is preferably, hydroxyl, cyano, alkoxy, =O, =S, NO<sub>2</sub>, halogen, N(CH<sub>3</sub>)<sub>2</sub>, amino, or SH. The term "alkyl" also includes alkynyl groups which have an unsaturated hydrocarbon group containing at least one carbon-carbon triple bond, including straight-chain, branched-chain, and cyclic groups. Preferably, the alkynyl group has 1 to 12 carbons. More preferably it is a lower alkynyl of from 1 to 7 carbons, more preferably 1 to 4 carbons. The alkynyl group may be substituted or unsubstituted. When

substituted the substituted group(s) is preferably, hydroxyl, cyano, alkoxy, =O, =S, NO<sub>2</sub> or N(CH<sub>3</sub>)<sub>2</sub>, amino or SH.

Such alkyl groups may also include aryl, alkylaryl, carbocyclic aryl, heterocyclic aryl, amide and ester groups. An "aryl" group refers to an aromatic group which has at least one ring having a conjugated p electron system and includes carbocyclic aryl, heterocyclic aryl and biaryl groups, all of which may be optionally substituted. The preferred substituent(s) of aryl groups are halogen, trihalomethyl, hydroxyl, SH, OH, cyano, alkoxy, alkyl, alkenyl, alkynyl, and amino groups. An "alkylaryl" group refers to an alkyl group (as described above) covalently joined to an aryl group (as described above).

5 Carbocyclic aryl groups are groups wherein the ring atoms on the aromatic ring are all carbon atoms. The carbon atoms are optionally substituted. Heterocyclic aryl groups are groups having from 1 to 3 heteroatoms as ring atoms in the aromatic ring and the remainder of the ring atoms are carbon atoms. Suitable heteroatoms include oxygen, sulfur, and nitrogen, and include furanyl, thienyl, pyridyl, pyrrolyl, N-lower alkyl pyrrolo,

10 pyrimidyl, pyrazinyl, imidazolyl and the like, all optionally substituted. An "amide" refers to an -C(O)-NH-R, where R is either alkyl, aryl, alkylaryl or hydrogen. An "ester" refers to an -C(O)-OR', where R is either alkyl, aryl, alkylaryl or hydrogen.

15

A "blocking group" is a group which is able to be removed after polynucleotide synthesis and/or which is compatible with solid phase polynucleotide synthesis.

20 A "phosphorus containing group" can include phosphorus in forms such as dithioates, phosphoramidites and/or as part of an oligonucleotide.

In yet another preferred embodiment C' is selected from a group comprising inverted abasic residue, 4',5'-methylene nucleotide; 1-(beta-D-erythofuranosyl) nucleotide, 4'-thio nucleotide, carbocyclic nucleotide; 1,5-anhydrohexitol nucleotide; L-25 nucleotides; alpha-nucleotides; modified base nucleotide; phosphorodithioate linkage; *threo*-pentofuranosyl nucleotide; acyclic 3',4'-seco nucleotide; acyclic 3,4-dihydroxybutyl nucleotide; acyclic 3,5-dihydroxypentyl nucleotide, 3'-3'-inverted nucleotide moiety; 3'-3'-inverted abasic moiety; 3'-2'-inverted nucleotide moiety; 3'-2'-inverted abasic moiety; 1,4-30 butanediol phosphate; 3'-phosphoramidate; hexylphosphate; aminohexyl phosphate; 3'-phosphate; 3'-phosphorothioate; phosphorodithioate; or bridging or non-bridging methylphosphonate moiety (for more details see Beigelman *et al.*, International PCT publication No. WO 97/26270, incorporated by reference herein).

In yet another preferred embodiment C is selected from a group comprising, 4',5'-methylene nucleotide; 1-(beta-D-erythofuranosyl) nucleotide; 4'-thio nucleotide, 35 carbocyclic nucleotide; 5'-amino-alkyl phosphate; 1,3-diamino-2-propyl phosphate, 3-aminopropyl phosphate; 6-aminohexyl phosphate; 1,2-aminododecyl phosphate;

hydroxypropyl phosphate; 1,5-anhydrohexitol nucleotide; L-nucleotide; alpha-nucleotide; modified base nucleotide; phosphorodithioate; *threo*-pentofuranosyl nucleotide; acyclic 3',4'-seco nucleotide; 3,4-dihydroxybutyl nucleotide; 3,5-dihydroxypentyl nucleotide, 5'-5'-inverted nucleotide moiety; 5'-5'-inverted abasic moiety; 5'-phosphoramidate; 5'-phosphorothioate; 1,4-butanediol phosphate; 5'-amino; bridging and/or non-bridging 5'-phosphoramidate, phosphorothioate and/or phosphorodithioate, bridging or non bridging methylphosphonate and 5'-mercapto moieties (for more details see Beaucage and Iyer, 1993, *Tetrahedron* 49, 1925; incorporated by reference herein).

In another preferred embodiment (Z)<sub>p</sub> includes a non-nucleotide linker. Thus, in a preferred embodiment, the invention features an enzymatic nucleic acid molecule having one or more non-nucleotide moieties, and having enzymatic activity to cleave an RNA or DNA molecule. By the term "non-nucleotide" is meant any group or compound which can be incorporated into a nucleic acid chain in the place of one or more nucleotide units, including either sugar and/or phosphate substitutions, and allows the remaining bases to exhibit their enzymatic activity. The group or compound is abasic in that it does not contain a commonly recognized nucleotide base, such as adenosine, guanine, cytosine, uracil or thymine. The terms "abasic" or "abasic nucleotide" as used herein encompass sugar moieties lacking a base or having other chemical groups in place of base at the 1' position.

By the phrase "enzymatic nucleic acid" is meant a nucleic acid molecule capable of catalyzing (altering the velocity and/or rate of) a variety of reactions including the ability to repeatedly cleave other separate nucleic acid molecules (endonuclease activity) in a nucleotide base sequence-specific manner. Such a molecule with endonuclease activity may have complementarity in a substrate binding region (e.g. (X)<sub>m</sub>, (X)<sub>o</sub>, (X)<sub>q</sub> and (Y)<sub>n</sub> in formulae IV-VII) to a specified gene target, and also has an enzymatic activity that specifically cleaves RNA or DNA in that target. That is, the nucleic acid molecule with endonuclease activity is able to intramolecularly or intermolecularly cleave RNA or DNA and thereby inactivate a target RNA or DNA molecule. This complementarity functions to allow sufficient hybridization of the enzymatic RNA molecule to the target RNA or DNA to allow the cleavage to occur. 100% complementarity is preferred, but complementarity as low as 50-75% may also be useful in this invention. The nucleic acids may be modified at the base, sugar, and/or phosphate groups. The term enzymatic nucleic acid is used interchangeably with phrases such as ribozymes, catalytic RNA, enzymatic RNA, catalytic DNA, nucleozyme, DNAzyme, RNA enzyme, endoribonuclease, endonuclease, minizyme, leadzyme, oligozyme, chimeric ribozyme, chimeric enzymatic nucleic acid, or

DNA enzyme. All of these terminologies describe nucleic acid molecules with enzymatic activity.

By "complementarity" is meant a nucleic acid that can form hydrogen bond(s) with other RNA sequence by either traditional Watson-Crick or other non-traditional types (for 5 example, Hoogsteen type) of base-paired interactions.

By "oligonucleotide" as used herein, is meant a molecule comprising two or more nucleotides.

By "enzymatic portion" is meant that part of the enzymatic nucleic acid molecule essential for cleavage of a nucleic acid substrate

10 By "substrate binding region" or "substrate binding domain" is meant that portion/region of a nucleic acid molecule (e.g. ribozyme) which is complementary to (*i.e.*, able to base-pair with) a portion of its substrate. Generally, such complementarity is 100%, but can be less if desired. For example, as few as 10 bases out of 14 may be base-paired. Such arms are shown generally in Figures 1 and 3. That is, in a ribozyme example, 15 these arms contain sequences within a ribozyme which are intended to bring ribozyme and target RNA together through complementary base-pairing interactions. The ribozyme of the invention may have binding arms that are contiguous or non-contiguous and may be of varying lengths. The length of the binding arm(s) are preferably greater than or equal to four nucleotides; specifically 12-100 nucleotides; more specifically 14-24 nucleotides 20 long. If two binding arms are chosen, the design is such that the length of the binding arms are symmetrical (*i.e.*, each of the binding arms is of the same length; *e.g.*, five and five nucleotides, six and six nucleotides or seven and seven nucleotides long) or asymmetrical (*i.e.*, the binding arms are of different length; *e.g.*, six and three nucleotides; three and six nucleotides long; four and five nucleotides long; four and six nucleotides 25 long; four and seven nucleotides long; and the like).

By "DNAzyme" or "catalytic DNA" or "DNA enzyme" is meant, an enzymatic nucleic acid molecule lacking a 2'-OH group.

By "nucleic acid molecule" as used herein is meant a molecule comprising nucleotides. The nucleic acid can be composed of modified or unmodified nucleotides or 30 non-nucleotides or various mixtures and combinations thereof.

In another preferred embodiment, the nucleic acid molecule of the present invention is conjugated with another moiety including but not limited to abasic nucleotides, polyether, polyamine, polyamides, peptides, carbohydrates, lipid, or polyhydrocarbon compounds. Those skilled in the art will recognize that these molecules 35 may be linked to one or more of any nucleotides comprising the nucleic acid molecule at several positions on the sugar, base or phosphate group.

In yet another preferred embodiment, the nucleic acid molecule of the present invention can form structures including but not limited to antisense, triplexes, 2-5A chimera antisense, or enzymatic nucleic acid (ribozymes).

By "antisense" is meant a non-enzymatic nucleic acid molecule that binds to target RNA, for example, by means of RNA-RNA or RNA-DNA or RNA-PNA (protein nucleic acid; Egholm *et al.*, 1993 *Nature* 365, 566) interactions and alters the activity of the target RNA (for a review see Stein and Cheng, 1993 *Science* 261, 1004).

By "2-5A antisense chimera" it is meant, an antisense oligonucleotide containing a 5' phosphorylated 2'-5'-linked adenylate residues. These chimeras bind to target RNA in a sequence-specific manner and activate a cellular 2-5A-dependent ribonuclease which in turn cleaves the target RNA (Torrence *et al.*, 1993 *Proc. Natl. Acad. Sci. USA* 90, 1300).

By "triplex DNA" it is meant an oligonucleotide that can bind to a double-stranded DNA in a sequence-specific manner to form a triple-strand helix. Triple-helix formation has been shown to inhibit transcription of the targeted gene (Duval-Valentin *et al.*, 1992 *Proc. Natl. Acad. Sci. USA* 89, 504).

In another preferred embodiment, the invention features an antisense oligonucleotide which is capable of interacting with the target RNA and sterically blocking translation, where the oligonucleotide has a 5' and a 3' Cap structure and the oligonucleotide may include modifications at the base, sugar or the phosphate groups.

Other features and advantages of the invention will be apparent from the following description of the preferred embodiments thereof, and from the claims.

#### Description of the Preferred Embodiments

The drawings will first briefly be described.

#### Drawings:

Figure 1 is a diagrammatic representation of a nucleic acid molecule with 7-9 phosphorothioate oligodeoxyribonucleotide sequence flanked by 9 non-deoxyribonucleotide containing oligonucleotides, binding to a target molecule.

Figure 2 displays schematic representations of certain chemical structures which may be incorporated into the nucleic acid molecule of the invention.

Figure 3 shows the secondary structure model for seven different classes of enzymatic nucleic acid molecules. Arrow indicates the site of cleavage. ----- indicate the target sequence. Lines interspersed with dots are meant to indicate tertiary interactions. - is meant to indicate base-paired interaction. **Group I Intron:** P1-P9.0 represent various stem-loop structures (Cech *et al.*, 1994, *Nature Struc. Bio.*, 1, 273). **RNase P (M1RNA):**

EGS represents external guide sequence (Forster *et al.*, 1990, *Science*, 249, 783; Pace *et al.*, 1990, *J. Biol. Chem.*, 265, 3587). **Group II Intron:** 5'SS means 5' splice site; 3' SS means 3'-splice site; IBS means intron binding site; EBS means exon binding site (Pyle *et al.*, 1994, *Biochemistry*, 33, 2716). **VS RNA:** I-VI are meant to indicate six stem-loop structures; shaded regions are meant to indicate tertiary interaction (Collins, International PCT Publication No. WO 96/19577). **HDV Ribozyme:** I-IV are meant to indicate four stem-loop structures (Been *et al.*, US Patent No. 5,625,047). **Hammerhead Ribozyme:** I-III are meant to indicate three stem-loop structures; stems I-III can be of any length and may be symmetrical or asymmetrical (Usman *et al.*, 1996, *Curr. Op. Struct. Bio.*, 1, 527).

**Hairpin Ribozyme:** Helix 1, 4 and 5 can be of any length; Helix 2 is between 3 and 8 base-pairs long; Y is a pyrimidine; Helix 2 (H2) is provided with a least 4 base pairs (*i.e.*, n is 1, 2, 3 or 4) and helix 5 can be optionally provided of length 2 or more bases (preferably 3 - 20 bases, *i.e.*, m is from 1 - 20 or more). Helix 2 and helix 5 may be covalently linked by one or more bases (*i.e.*, r is 1 base). Helix 1, 4 or 5 may also be extended by 2 or more base pairs (*e.g.*, 4 - 20 base pairs) to stabilize the ribozyme structure, and preferably is a protein binding site. In each instance, each N and N' independently is any normal or modified base and each dash represents a potential base-pairing interaction. These nucleotides may be modified at the sugar, base or phosphate. Complete base-pairing is not required in the helices, but is preferred. Helix 1 and 4 can be of any size (*i.e.*, o and p is each independently from 0 to any number, *e.g.*, 20) as long as some base-pairing is maintained. Essential bases are shown as specific bases in the structure, but those in the art will recognize that one or more may be modified chemically (abasic, base, sugar and/or phosphate modifications) or replaced with another base without significant effect. Helix 4 can be formed from two separate molecules, *i.e.*, without a connecting loop. The connecting loop when present may be a ribonucleotide with or without modifications to its base, sugar or phosphate. "q" is 2 bases. The connecting loop can also be replaced with a non-nucleotide linker molecule. H refers to bases A, U, or C. Y refers to pyrimidine bases. "—" refers to a covalent bond. (Burke *et al.*, 1996, *Nucleic Acids & Mol. Biol.*, 10, 129; Chowrira *et al.*, US Patent No. 5,631,359).

Figure 4 is a graph comparing cell proliferation rates of MCF-7 cells treatment with active and inactive ribozyme with mismatch arms targeted to estrogen receptor delivered with GSV transfection reagent. The sequence for the active ribozyme is given as Seq. ID. No.2725 and the inactive is 2726.

Figure 5 is a graph comparing RNA levels of c-raf RNA in PC-3 cells following treatment with antisense (Seq. ID. No 2717) and scrambled antisense (mismatch) controls.

Figure 6a and 6b displays several possible ribozymes comprising oligodeoxyribonucleotides. The symbols used in the diagram include: N' represents a nucleotide complementary to a nucleotide on the target molecule; N7 represents position 7 in the ribozyme molecule (Hertel, K. J., *et al.*, 1992, *Nucleic Acids Res.*, 20, 3252); N' represents a deoxyribonucleotide complementary to a nucleotide on the target molecule; s represents a phosphorothioate modification; C represents a chemical modification at the 5' end of the ribozyme; and C' represents a chemical modification at the 3' end.

Figure 7 shows examples of nucleotide modifications for incorporation into oligonucleotides.

Figure 8 is a graph demonstrating the level of c-raf mRNA in PC-3 cells following treatment with c-raf sequence targeting antisense nucleic acid molecules. The antisense molecules target regions within the intron/exon junction (Seq. I.D. Nos. 2731-2735), intron (Seq. I.D. 2736) and exon (Seq. I.D. 2737). The results demonstrate the molecules' ability to inhibit c-raf mRNA compared to the untreated or mismatch control at two concentrations.

Figure 9 is a graph which demonstrates the ability of an antisense molecule represented by Seq. I.D. No. 2738 (phosphorothioate modifications at every deoxynucleotide position) to inhibit c-raf message in PC-3 cells over a period of five days compared to the untreated and mismatch controls at three different concentrations.

Figure 10 is a graph which demonstrates the ability of an antisense molecule represented by Seq. I.D. No. 2737 (three phosphorothioate modified nucleotides at both the 5' and 3' ends of the oligonucleotide) to inhibit c-raf message in PC-3 cells over a period of five days compared to the untreated control at three different concentrations.

Figure 11 is a graph which demonstrates the ability of an antisense molecule represented by Seq. I.D. No. 2744 (9 phosphorothioate modified DNA flanked by 7 2'-O-methylthiomethyl RNA nucleotides at the 5' and 3' end) to inhibit c-raf message in PC-3 cells over a period of five days compared to the untreated and mismatch control at three different concentrations.

Figure 12 is a graph showing the level of cellular proliferation inhibition exhibited by antisense oligonucleotides represented by Seq. I.D. Nos. 2738 and 2741. Also shown within the graph is the cellular proliferation after treatment of cells with a mismatch control (Seq. I.D. 2739).

Figure 13 is a graph showing the ability of oligonucleotides with different chemical modifications to inhibit c-raf mRNA. Each antisense molecule is compared to an untreated and mismatch control.

Figure 14 is a graph demonstrating a dose dependent inhibition of C-raf in PC-3 cells following treatment with antisense oligonucleotides (Seq. I.D. Nos. 2741 and 2738).

Figure 15 is a graph showing inhibition of bcl-2 mRNA by antisense oligonucleotides compared to untreated and mismatch controls.

5 Figure 16 is a graph showing the ability of several k-ras targeting antisense molecules to inhibit k-ras message

Figure 17 is a graph showing the ability of oligonucleotides with different chemical modifications to inhibit estrogen receptor mRNA. Each antisense molecule is compared to an untreated and mismatch control.

10 Figure 18 shows a scheme for the synthesis of 3'-deoxy-3'-thio guanosine nucleoside (scheme 1).

Figure 19 shows a scheme for the synthesis of S-(pyridyl-2-disulfanyl) derivative (scheme 2).

15 Figure 20 shows a scheme for the synthesis of 3'-deoxy-3'-thio guanosine phosphoramide.

Figure 21 shows a scheme for the preparation of 5'-thio-nucleoside phosphoramide and succinates.

Figure 22 shows a scheme for the synthesis of 3'-thio-2'-O-methyl uridine.

#### Synthesis of Nucleic acid Molecules

20 Synthesis of nucleic acids greater than 100 nucleotides in length is difficult using automated methods, and the therapeutic cost of such molecules is prohibitive. In this invention, small nucleic acid motifs (e.g., antisense oligonucleotides, hammerhead or the hairpin ribozymes) are used for exogenous delivery. The simple structure of these molecules increases the ability of the nucleic acid to invade targeted regions of RNA structure. The molecules of the instant invention were chemically synthesized.

25 Oligodeoxyribonucleotides were synthesized using standard protocols as described in Caruthers *et al.*, 1992, *Methods in Enzymology* 211,3-19, and is incorporated by reference.

The method of synthesis used for normal RNA including certain enzymatic nucleic acid molecules follows the procedure as described in Usman *et al.*, 1987 *J. Am. Chem. Soc.*, 109, 7845; Scaringe *et al.*, 1990 *Nucleic Acids Res.*, 18, 5433; and Wincott *et al.*, 1995 *Nucleic Acids Res.* 23, 2677-2684 and makes use of common nucleic acid protecting and coupling groups, such as dimethoxytrityl at the 5'-end, and phosphoramidites at the 3'-end. In a non-limiting example, small scale synthesis were conducted on a 394 Applied Biosystems, Inc. synthesizer using a modified 2.5  $\mu$ mol scale protocol with a 5 min 35 coupling step for alkylsilyl protected nucleotides and 2.5 min coupling step for 2'-O-

methylated nucleotides. Table **II** outlines the amounts, and the contact times, of the reagents used in the synthesis cycle. A 6.5-fold excess (163  $\mu$ L of 0.1 M = 16.3  $\mu$ mol) of phosphoramidite and a 24-fold excess of *S*-ethyl tetrazole (238  $\mu$ L of 0.25 M = 59.5  $\mu$ mol) relative to polymer-bound 5'-hydroxyl was used in each coupling cycle. Average coupling 5 yields on the 394 Applied Biosystems, Inc. synthesizer, determined by colorimetric quantitation of the trityl fractions, were 97.5-99%. Other oligonucleotide synthesis reagents for the 394 Applied Biosystems, Inc. synthesizer, detritylation solution was 2% TCA in methylene chloride (ABI); capping was performed with 16% *N*-methyl imidazole in THF (ABI) and 10% acetic anhydride/10% 2,6-lutidine in THF (ABI); oxidation 10 solution was 16.9 mM  $I_2$ , 49 mM pyridine, 9% water in THF (Millipore). B & J Synthesis Grade acetonitrile was used directly from the reagent bottle. *S*-Ethyl tetrazole solution (0.25 M in acetonitrile) was made up from the solid obtained from American International Chemical, Inc.

Deprotection of the RNA was performed as follows. The polymer-bound 15 oligoribonucleotide, trityl-off, was transferred from the synthesis column to a 4mL glass screw top vial and suspended in a solution of methylamine (MA) at 65 °C for 10 min. After cooling to -20 °C, the supernatant was removed from the polymer support. The support was washed three times with 1.0 mL of EtOH:MeCN:H<sub>2</sub>O/3:1:1, vortexed and the supernatant was then added to the first supernatant. The combined supernatants, 20 containing the oligoribonucleotide, were dried to a white powder.

The base-deprotected oligoribonucleotide was resuspended in anhydrous TEA•HF/NMP solution (250  $\mu$ L of a solution of 1.5mL *N*-methylpyrrolidinone, 750  $\mu$ L TEA and 1.0 mL TEA•3HF to provide a 1.4M HF concentration) and heated to 65°C for 1.5 h. The resulting, fully deprotected, oligomer was quenched with 50 mM TEAB (9 25 mL) prior to anion exchange desalting.

For anion exchange desalting of the deprotected oligomer, the TEAB solution was loaded onto a Qiagen 500® anion exchange cartridge (Qiagen Inc.) that was prewashed with 50 mM TEAB (10 mL). After washing the loaded cartridge with 50 mM TEAB (10 mL), the RNA was eluted with 2 M TEAB (10 mL) and dried down to a white powder.

30 Inactive hammerhead ribozymes were synthesized by substituting a U for G5 and a U for A14 (numbering from Hertel, K. J., *et al.*, 1992, *Nucleic Acids Res.*, 20, 3252).

The average stepwise coupling yields were >98% (Wincott *et al.*, 1995 *Nucleic Acids Res.* 23, 2677-2684).

35 Alternatively, the nucleic acid molecules of the present invention can be synthesized separately and joined together by ligation (Moore *et al.*, 1992, *Science* 256,

9923; Draper *et al.*, International PCT publication No. WO 93/23569; Shabarova *et al.*, 1991, *Nucleic Acids Research* 19, 4247)

Administration of Nucleic Acid Molecules

Methods for the delivery of nucleic acid molecules is described in Akhtar *et al.*, 5 1992, *Trends Cell Bio.*, 2, 139; and *Delivery Strategies for Antisense Oligonucleotide Therapeutics*, ed. Akhtar, 1995 which are both incorporated herein by reference. Sullivan *et al.*, PCT WO 94/02595, further describes the general methods for delivery of enzymatic RNA molecules. These protocols may be utilized for the delivery of virtually any nucleic acid molecule. Nucleic acid molecules may be administered to cells by a variety of 10 methods known to those familiar to the art, including, but not restricted to, encapsulation in liposomes, by iontophoresis, or by incorporation into other vehicles, such as hydrogels, cyclodextrins, biodegradable nanocapsules, and bioadhesive microspheres. For some indications, nucleic acid molecules may be directly delivered *ex vivo* to cells or tissues with or without the aforementioned vehicles. Alternatively, the nucleic acid/vehicle 15 combination is locally delivered by direct injection or by use of a catheter, infusion pump or stent. Other routes of delivery include, but are not limited to, intravascular, intramuscular, subcutaneous or joint injection, aerosol inhalation, oral (tablet or pill form), topical, systemic, ocular, intraperitoneal and/or intrathecal delivery. More detailed descriptions of nucleic acid delivery and administration are provided in Sullivan *et al.*, 20 supra and Draper *et al.*, PCT WO93/23569 which have been incorporated by reference herein.

The molecules of the instant invention can be used as pharmaceutical agents. Pharmaceutical agents prevent, inhibit the occurrence, or treat (alleviate a symptom to some extent, preferably all of the symptoms) of a disease state in a patient.

25 The negatively charged polynucleotides of the invention can be administered (e.g., RNA, DNA or protein) and introduced into a patient by any standard means, with or without stabilizers, buffers, and the like, to form a pharmaceutical composition. When it is desired to use a liposome delivery mechanism, standard protocols for formation of liposomes can be followed. The compositions of the present invention may also be 30 formulated and used as tablets, capsules or elixirs for oral administration; suppositories for rectal administration; sterile solutions; suspensions for injectable administration; and the like.

The present invention also includes pharmaceutically acceptable formulations of the compounds described. These formulations include salts of the above compounds, e.g.,

acid addition salts, for example, salts of hydrochloric, hydrobromic, acetic acid, and benzene sulfonic acid.

A pharmacological composition or formulation refers to a composition or formulation in a form suitable for administration, *e.g.*, systemic administration, into a cell 5 or patient, preferably a human. Suitable forms, in part, depend upon the use or the route of entry, for example oral, transdermal, or by injection. Such forms should not prevent the composition or formulation to reach a target cell (*i.e.*, a cell to which the negatively charged polymer is desired to be delivered to). For example, pharmacological compositions injected into the blood stream should be soluble. Other factors are known in 10 the art, and include considerations such as toxicity and forms which prevent the composition or formulation from exerting its effect.

By "systemic administration" is meant *in vivo* systemic absorption or accumulation of drugs in the blood stream followed by distribution throughout the entire body. Administration routes which lead to systemic absorption include, without 15 limitations: intravenous, subcutaneous, intraperitoneal, inhalation, oral, intrapulmonary and intramuscular. Each of these administration routes expose the desired negatively charged polymers, *e.g.*, nucleic acids, to an accessible diseased tissue. The rate of entry of a drug into the circulation has been shown to be a function of molecular weight or size. The use of a liposome or other drug carrier comprising the compounds of the instant 20 invention can potentially localize the drug, for example, in certain tissue types, such as the tissues of the reticular endothelial system (RES). A liposome formulation which can facilitate the association of drug with the surface of cells, such as, lymphocytes and macrophages is also useful. This approach may provide enhanced delivery of the drug to target cells by taking advantage of the specificity of macrophage and lymphocyte immune 25 recognition of abnormal cells, such as the cancer cells.

The invention also features the use of the a composition comprising surface-modified liposomes containing poly (ethylene glycol) lipids (PEG-modified, or long-circulating liposomes or stealth liposomes). These formulations offer an method for increasing the accumulation of drugs in target tissues. This class of drug carriers resists 30 opsonization and elimination by the mononuclear phagocytic system (MPS or RES), thereby enabling longer blood circulation times and enhanced tissue exposure for the encapsulated drug (Lasic *et al.* *Chem. Rev.* 1995, **95**, 2601-2627; Ishiwata *et al.*, *Chem. Pharm. Bull.* 1995, **43**, 1005-1011). Such liposomes have been shown to accumulate 35 selectively in tumors, presumably by extravasation and capture in the neovascularized target tissues (Lasic *et al.*, *Science* 1995, **267**, 1275-1276; Oku *et al.*, 1995, *Biochim. Biophys. Acta*, **1238**, 86-90). The long-circulating liposomes enhance the

pharmacokinetics and pharmacodynamics of DNA and RNA, particularly compared to conventional cationic liposomes which are known to accumulate in tissues of the MPS (Liu *et al.*, *J. Biol. Chem.* 1995, **42**, 24864-24870; Choi *et al.*, International PCT Publication No. WO 96/10391; Ansell *et al.*, International PCT Publication No. WO 5 96/10390; Holland *et al.*, International PCT Publication No. WO 96/10392; all of these are incorporated by reference herein). Long-circulating liposomes are also likely to protect drugs from nuclease degradation to a greater extent compared to cationic liposomes, based on their ability to avoid accumulation in metabolically aggressive MPS tissues such as the liver and spleen. All of these references are incorporated by reference 10 herein.

10 The present invention also includes compositions prepared for storage or administration which include a pharmaceutically effective amount of the desired compounds in a pharmaceutically acceptable carrier or diluent. Acceptable carriers or diluents for therapeutic use are well known in the pharmaceutical art, and are described, 15 for example, in *Remington's Pharmaceutical Sciences*, Mack Publishing Co. (A.R. Gennaro edit. 1985) hereby incorporated by reference herein. For example, preservatives, stabilizers, dyes and flavoring agents may be provided. *Id.* at 1449. These include sodium benzoate, sorbic acid and esters of *p*-hydroxybenzoic acid. In addition, antioxidants and suspending agents may be used. *Id.*

20 A pharmaceutically effective dose is that dose required to prevent, inhibit the occurrence, or treat (alleviate a symptom to some extent, preferably all of the symptoms) of a disease state. The pharmaceutically effective dose depends on the type of disease, the composition used, the route of administration, the type of mammal being treated, the physical characteristics of the specific mammal under consideration, concurrent 25 medication, and other factors which those skilled in the medical arts will recognize. Generally, an amount between 0.1 mg/kg and 100 mg/kg body weight/day of active ingredients is administered dependent upon potency of the negatively charged polymer.

#### Mechanism of action of Nucleic Acid Molecules of the Invention

30 Antisense: Antisense molecules may be RNA or DNA oligonucleotides and primarily function by specifically binding to matching sequences resulting in inhibition of peptide synthesis (Wu-Pong, Nov 1994, *BioPharm*, 20-33). The oligonucleotide binds to target RNA by Watson Crick base-pairing and blocks gene expression by preventing 35 ribosomal translation of the bound sequences. Antisense molecules may also alter protein synthesis by interfering with RNA processing or transport from the nucleus into the cytoplasm (Mukhopadhyay & Roth, 1996, *Crit. Rev. in Oncogenesis* 7, 151-190).

In addition, binding of single stranded DNA to RNA may result in nuclease degradation of the heteroduplex (Wu-Pong, *supra*; Crooke, *supra*). To date, the only backbone modified DNA chemistry which will act as substrates for RNase H are phosphorothioates and phosphorodithioates. In experiments with *E. coli*, the 5 oligodeoxyribonucleotides phosphorothioate modification activated RNase H more efficiently (2-5 fold) compared to the natural phosphodiester containing oligodeoxynucleotide (Crooke, 1995, *supra*). Applicant describes here for the first time that oligonucleotides with 5'-thiophosphate modification can activate RNase H cleavage of RNA.

10 Triplex Forming Oligonucleotides (TFO): Single stranded DNA may be designed to bind to genomic DNA in a sequence specific manner. TFOs are comprised of pyrimidine-rich oligonucleotides which bind DNA helices through Hoogsteen Base-pairing (Wu-Pong, *supra*). The resulting triple helix composed of the DNA sense, DNA antisense, and TFO disrupts RNA synthesis by RNA polymerase. The TFO mechanism 15 may result in gene expression or cell death since binding may be irreversible (Mukhopadhyay & Roth, *supra*).

20 2-5A Antisense Chimera: The 2-5A system is an interferon mediated mechanism for RNA degradation found in higher vertebrates (Mitra *et al.*, 1996, *Proc Nat Acad Sci USA* 93, 6780-6785). Two types of enzymes, 2-5A synthetase and RNase L, are required 25 for RNA cleavage. The 2-5A synthetases require double stranded RNA to form 2'-5' oligoadenylylates (2-5A). 2-5A then acts as an allosteric effector for utilizing RNase L which has the ability to cleave single stranded RNA. The ability to form 2-5A structures with double stranded RNA makes this system particularly useful for inhibition of viral replication.

25 (2'-5')oligoadenylate structures may be covalently linked to antisense molecules to form chimeric oligonucleotides capable of RNA cleavage (Torrence, *supra*). These molecules putatively bind and activate a 2-5A dependent RNase, the oligonucleotide/enzyme complex then binds to a target RNA molecule which can then be cleaved by the RNase enzyme.

30 Enzymatic Nucleic acid: Seven basic varieties of naturally-occurring enzymatic RNAs are known presently. In addition, several *in vitro* selection (evolution) strategies (Orgel, 1979, *Proc. R. Soc. London*, B 205, 435) have been used to evolve new nucleic acid catalysts capable of catalyzing cleavage and ligation of phosphodiester linkages (Joyce, 1989, *Gene*, 82, 83-87; Beaudry *et al.*, 1992, *Science* 257, 635-641; Joyce, 1992, 35 *Scientific American* 267, 90-97; Breaker *et al.*, 1994, *TIBTECH* 12, 268; Bartel *et al.*, 1993, *Science* 261:1411-1418; Szostak, 1993, *TIBS* 17, 89-93; Kumar *et al.*, 1995, *FASEB J.*, 9,

1183; Breaker, 1996, *Curr. Op. Biotech.*, 7, 442; Santoro *et al.*, 1997, *Proc. Natl. Acad. Sci.*, 94, 4262; Tang *et al.*, 1997, *RNA* 3, 914; Nakamaye & Eckstein, 1994, *supra*; Long & Uhlenbeck, 1994, *supra*; Ishizaka *et al.*, 1995, *supra*; Vaish *et al.*, 1997, *Biochemistry* 36, 6495; all of these are incorporated by reference herein). Each can catalyze a series of 5 reactions including the hydrolysis of phosphodiester bonds in *trans* (and thus can cleave other RNA molecules) under physiological conditions. Table I summarizes some of the characteristics of some of these ribozymes. In general, enzymatic nucleic acids act by first binding to a target RNA. Such binding occurs through the target binding portion of an enzymatic nucleic acid which is held in close proximity to an enzymatic portion of the 10 molecule that acts to cleave the target RNA. Thus, the enzymatic nucleic acid first recognizes and then binds a target RNA through complementary base-pairing, and once bound to the correct site, acts enzymatically to cut the target RNA. Strategic cleavage of such a target RNA will destroy its ability to direct synthesis of an encoded protein. After an enzymatic nucleic acid has bound and cleaved its RNA target, it is released from that 15 RNA to search for another target and can repeatedly bind and cleave new targets.

The enzymatic nature of a ribozyme is advantageous over other technologies, since the concentration of ribozyme necessary to affect a therapeutic treatment is lower. This advantage reflects the ability of the ribozyme to act enzymatically. Thus, a single ribozyme molecule is able to cleave many molecules of target RNA. In addition, the 20 ribozyme is a highly specific inhibitor, with the specificity of inhibition depending not only on the base-pairing mechanism of binding to the target RNA, but also on the mechanism of target RNA cleavage. Single mismatches, or base-substitutions, near the site of cleavage can be chosen to completely eliminate catalytic activity of a ribozyme.

Nucleic acid molecules having an endonuclease enzymatic activity are able to 25 repeatedly cleave other separate RNA molecules in a nucleotide base sequence-specific manner. Such enzymatic nucleic acid molecules can be targeted to virtually any RNA transcript, and efficient cleavage achieved *in vitro* (Zaug *et al.*, 324, *Nature* 429 1986; Uhlenbeck, 1987 *Nature* 328, 596; Kim *et al.*, 84 *Proc. Natl. Acad. Sci. USA* 8788, 1987; Dreyfus, 1988, *Einstein Quart. J. Bio. Med.*, 6, 92; Haseloff and Gerlach, 334 *Nature* 585, 30 1988; Cech, 260 *JAMA* 3030, 1988; and Jefferies *et al.*, 17 *Nucleic Acids Research* 1371, 1989; Santoro *et al.*, 1997 *supra*).

Because of their sequence-specificity, *trans*-cleaving ribozymes show promise as therapeutic agents for human disease (Usman & McSwiggen, 1995 *Ann. Rep. Med. Chem.* 30, 285-294; Christoffersen and Marr, 1995 *J. Med. Chem.* 38, 2023-2037). Ribozymes 35 can be designed to cleave specific RNA targets within the background of cellular RNA. Such a cleavage event renders the RNA non-functional and abrogates protein expression

from that RNA. In this manner, synthesis of a protein associated with a disease state can be selectively inhibited.

#### Optimizing Ribozyme Activity

Catalytic activity of the ribozymes described in the instant invention can be 5 optimized as described by Draper et al., *supra*. The details will not be repeated here, but include altering the length of the ribozyme binding arms, or chemically synthesizing ribozymes with modifications (base, sugar and/or phosphate) that prevent their degradation by serum ribonucleases and/or enhance their enzymatic activity (see e.g., Eckstein et al., International Publication No. WO 92/07065; Perrault et al., 1990 *Nature* 10 344, 565; Pieken et al., 1991 *Science* 253, 314; Usman and Cedergren, 1992 *Trends in Biochem. Sci.* 17, 334; Usman et al., International Publication No. WO 93/15187; and Rossi et al., International Publication No. WO 91/03162; Sproat, US Patent No. 15 5,334,711; and Burgin et al., *supra*; all of these describe various chemical modifications that can be made to the base, phosphate and/or sugar moieties of enzymatic RNA molecules). Modifications which enhance their efficacy in cells, and removal of bases from stem loop structures to shorten RNA synthesis times and reduce chemical requirements are desired. (All these publications are hereby incorporated by reference herein).

There are several examples in the art describing sugar, base and phosphate 20 modifications that can be introduced into enzymatic nucleic acid molecules without significantly effecting catalysis and with significant enhancement in their nuclease stability and efficacy. Ribozymes are modified to enhance stability and/or enhance catalytic activity by modification with nuclease resistant groups, for example, 2'-amino, 2'-C-allyl, 2'-flouro, 2'-O-methyl, 2'-H, nucleotide base modifications (for a review see 25 Usman and Cedergren, 1992 *TIBS* 17, 34; Usman et al., 1994 *Nucleic Acids Symp. Ser.* 31, 163; Burgin et al., 1996 *Biochemistry* 35, 14090). Sugar modification of enzymatic nucleic acid molecules have been extensively described in the art (see Eckstein et al., International Publication PCT No. WO 92/07065; Perrault et al. *Nature* 1990, 344, 565-568; Pieken et al. *Science* 1991, 253, 314-317; Usman and Cedergren, *Trends in Biochem. 30 Sci.* 1992, 17, 334-339; Usman et al. International Publication PCT No. WO 93/15187; Sproat, US Patent No. 5,334,711 and Beigelman et al., 1995 *J. Biol. Chem.* 270, 25702; all of the references are hereby incorporated in their totality by reference herein). Such 35 publications describe general methods and strategies to determine the location of incorporation of sugar, base and/or phosphate modifications and the like into ribozymes without inhibiting catalysis, and are incorporated by reference herein. In view of such

teachings, similar modifications can be used as described herein to modify the nucleic acid catalysts of the instant invention.

Nucleic acid catalysts having chemical modifications which maintain or enhance enzymatic activity are provided. Such nucleic acid is also generally more resistant to nucleases than unmodified nucleic acid. Thus, in a cell and/or *in vivo* the activity may not be significantly lowered. As exemplified herein such ribozymes are useful in a cell and/or *in vivo* even if activity over all is reduced 10 fold (Burgin *et al.*, 1996, *Biochemistry*, 35, 14090). Such ribozymes herein are said to "maintain" the enzymatic activity on all RNA ribozyme.

10 Therapeutic ribozymes delivered exogenously must optimally be stable within cells until translation of the target RNA has been inhibited long enough to reduce the levels of the undesirable protein. This period of time varies between hours to days depending upon the disease state. Clearly, ribozymes must be resistant to nucleases in order to function as effective intracellular therapeutic agents. Improvements in the chemical synthesis of RNA  
15 (Wincott *et al.*, 1995 *Nucleic Acids Res.* 23, 2677; incorporated by reference herein) have expanded the ability to modify ribozymes by introducing nucleotide modifications to enhance their nuclease stability as described above.

By "enhanced enzymatic activity" is meant to include activity measured in cells and/or *in vivo* where the activity is a reflection of both catalytic activity and ribozyme 20 stability. In this invention, the product of these properties is increased or not significantly (less than 10 fold) decreased *in vivo* compared to an all RNA ribozyme.

In yet another preferred embodiment, nucleic acid catalysts having chemical modifications which maintain or enhance enzymatic activity is provided. Such nucleic acid is also generally more resistant to nucleases than unmodified nucleic acid. Thus, in a 25 cell and/or *in vivo* the activity may not be significantly lowered. As exemplified herein such ribozymes are useful in a cell and/or *in vivo* even if activity over all is reduced 10 fold (Burgin *et al.*, 1996, *Biochemistry*, 35, 14090). Such ribozymes herein are said to "maintain" the enzymatic activity on all RNA ribozyme.

#### Inhibition Of Estrogen Receptor Gene Expression

30 Breast Cancer is one of the leading causes of death in women (Jiang and Jordan, 1992, *J. Natl. Cancer Inst.* 84, 580-591). There has been an intense effort to understand the molecular mechanisms for hormonal regulation of cell proliferation in breast cancer over the last several decades. It has been shown that many breast and endometrial cancers are dependent on estrogen for their growth and progression (Borras *et al.*, 1994, *J. Steroid  
35 Biochem. Molec. Biol.* 48, 325-336). Estrogen receptor plays a pivotal role in these cancers

and thus controlling the expression of this gene is of paramount interest to researchers and clinicians. The estrogen receptor is a member of the steroid hormone receptor gene family that displays its biological function as a ligand binding-dependent transcription factor. Tamoxifen is a nonsteroidal antiestrogen which treats all stages of breast cancer and may 5 be used as a preventative compound in those predisposed to breast cancer (Jordan and Murphy, 1990, *Endocr. Rev.* 11; 578-610).

Most breast tumors are initially dependent upon estrogen for growth, and the estrogen receptor has been a key indicator for endocrine response, prognosis and survival from breast cancer. The MCF-7 human breast cancer cell line expresses high levels of 10 estrogen receptor and is responsive to the effects of added estrogen (Borras *et al.*, 1996, *J. Steroid Biochem. Molec. Biol.* 57, 203-213; Pink and Jordan, 1996, *Cancer Res.* 56, 2321-2330). They are an excellent model system to study the effects of regulation of estrogen receptor in breast cancer. Ribozymes and antisense oligonucleotides represent a direct means of affecting the levels of estrogen receptor message. In estrogen dependent cell 15 lines, decreased amounts of estrogen receptor transcript should lower overall amounts of estrogen receptor protein and prevent proliferation of those cells. The effects of estrogen receptor on sexual differentiation of brain has been examined using antisense oligonucleotides (McCarthy *et al.*, 1993 *Endocrinology* 133, 433-439). This application documents the effects of ribozymes and antisense oligonucleotides to estrogen receptor 20 RNA levels and proliferation in MCF-7 cells.

Estrogen receptor may be inhibited using nucleic acid molecules, including the nucleic acid molecules of the present invention. Other references describe the use of antisense molecules to down regulate estrogen receptor RNA (Defazio *et al.*, 1997, *Cell Growth Differ.* 8, 903-911; Santagati *et al.*, 1997, *Mol. Endocrinol.* 11, 938-949; Williard 25 *et al.*, 1994, *Gene* 149, 21-24; Jiang & Jordan, *supra*).

The nucleic acid molecules may be chemically synthesized and delivered using methods described above, or may be expressed within cells from eukaryotic promoters (e.g., Izant and Weintraub, 1985 *Science* 229, 345; McGarry and Lindquist, 1986 *Proc. Natl. Acad. Sci. USA* 83, 399; Scanlon *et al.*, 1991, *Proc. Natl. Acad. Sci. USA*, 88, 10591-30 5; Kashani-Sabet *et al.*, 1992 *Antisense Res. Dev.*, 2, 3-15; Dropulic *et al.*, 1992 *J. Virol.* 66, 1432-41; Weerasinghe *et al.*, 1991 *J. Virol.*, 65, 5531-4; Ojwang *et al.*, 1992 *Proc. Natl. Acad. Sci. USA* 89, 10802-6; Chen *et al.*, 1992 *Nucleic Acids Res.*, 20, 4581-9; Sarver *et al.*, 1990 *Science* 247, 1222-1225; Thompson *et al.*, 1995 *Nucleic Acids Res.* 23, 2259; Good *et al.*, 1997, *Gene Therapy*, 4, 45; all of the references are hereby incorporated 35 in their totality by reference herein). Those skilled in the art realize that any nucleic acid can be expressed in eukaryotic cells from the appropriate DNA/RNA vector. The activity

of such nucleic acids can be augmented by their release from the primary transcript by a ribozyme (Draper *et al.*, PCT WO 93/23569, and Sullivan *et al.*, PCT WO 94/02595; Ohkawa *et al.*, 1992 *Nucleic Acids Symp. Ser.*, 27, 15-6; Taira *et al.*, 1991, *Nucleic Acids Res.*, 19, 5125-30; Ventura *et al.*, 1993 *Nucleic Acids Res.*, 21, 3249-55; Chowrira *et al.*, 5 1994 *J. Biol. Chem.* 269, 25856; all of the references are hereby incorporated in their totality by reference herein).

One type of nucleic acid molecules known as ribozymes, which can cleave target molecules, are expressed from transcription units (see for example Couture *et al.*, 1996, *TIG.*, 12, 510) inserted into DNA or RNA vectors. The recombinant vectors are preferably 10 DNA plasmids or viral vectors. Ribozyme expressing viral vectors could be constructed based on, but not limited to, adeno-associated virus, retrovirus, adenovirus, or alphavirus. Preferably, the recombinant vectors capable of expressing the ribozymes are delivered as described above, and persist in target cells. Alternatively, viral vectors may be used that provide for transient expression of ribozymes. Such vectors might be repeatedly 15 administered as necessary. Once expressed, the ribozymes cleave the target mRNA. The active ribozyme contains an enzymatic center or core equivalent to those in the examples, and binding arms able to bind target nucleic acid molecules such that cleavage at the target site occurs. Other sequences may be present which do not interfere with such cleavage. Delivery of ribozyme expressing vectors could be systemic, such as by 20 intravenous or intramuscular administration, by administration to target cells ex-planted from the patient followed by reintroduction into the patient, or by any other means that would allow for introduction into the desired target cell (for a review see Couture *et al.*, 1996, *TIG.*, 12, 510).

In one aspect the invention features, an expression vector comprising nucleic acid 25 sequence encoding at least one of the nucleic acid molecules of the instant invention is disclosed. The nucleic acid sequence encoding the nucleic acid catalyst of the instant invention is operably linked in a manner which allows expression of that nucleic acid molecule.

In another aspect the invention features, the expression vector comprises: a 30 transcription initiation region (e.g., eukaryotic pol I, II or III initiation region); b) a transcription termination region (e.g., eukaryotic pol I, II or III termination region); c) a gene encoding at least one of the nucleic acid catalyst of the instant invention; and wherein said gene is operably linked to said initiation region and said termination region, in a manner which allows expression and/or delivery of said nucleic acid molecule. The vector 35 may optionally include an open reading frame (ORF) for a protein operably linked on the

5' side or the 3'-side of the gene encoding the nucleic acid catalyst of the invention; and/or an intron (intervening sequences).

Transcription of the ribozyme sequences are driven from a promoter for eukaryotic RNA polymerase I (pol I), RNA polymerase II (pol II), or RNA polymerase III (pol III). 5 Transcripts from pol II or pol III promoters will be expressed at high levels in all cells; the levels of a given pol II promoter in a given cell type will depend on the nature of the gene regulatory sequences (enhancers, silencers, etc.) present nearby. Prokaryotic RNA polymerase promoters are also used, providing that the prokaryotic RNA polymerase enzyme is expressed in the appropriate cells (Elroy-Stein and Moss, 1990 *Proc. Natl. Acad. Sci. U S A*, 87, 6743-7; Gao and Huang 1993 *Nucleic Acids Res.*, 21, 2867-72; Lieber et al., 1993 *Methods Enzymol.*, 217, 47-66; Zhou et al., 1990 *Mol. Cell. Biol.*, 10, 4529-37). Several investigators have demonstrated that ribozymes expressed from such promoters can function in mammalian cells (e.g. Kashani-Sabet et al., 1992 *Antisense Res. Dev.*, 2, 3-15; Ojwang et al., 1992 *Proc. Natl. Acad. Sci. U S A*, 89, 10802-6; Chen et al., 1992 *Nucleic Acids Res.*, 20, 4581-9; Yu et al., 1993 *Proc. Natl. Acad. Sci. U S A*, 90, 6340-4; L'Huillier et al., 1992 *EMBO J.* 11, 4411-8; Lisziewicz et al., 1993 *Proc. Natl. Acad. Sci. U. S. A.*, 90, 8000-4; Thompson et al., 1995 *Nucleic Acids Res.* 23, 2259; Sullenger & Cech, 1993, *Science*, 262, 1566). More specifically, transcription units such as the ones derived from genes encoding U6 small nuclear (snRNA), transfer RNA (tRNA) and adenovirus VA RNA are useful in generating high concentrations of desired RNA molecules such as ribozymes in cells (Thompson et al., *supra*; Couture and Stinchcomb, 1996, *supra*; Noonberg et al., 1994, *Nucleic Acid Res.*, 22, 2830; Noonberg et al., US Patent No. 5,624,803; Good et al., 1997, *Gene Ther.* 4, 45; Beigelman et al., International PCT Publication No. WO 96/18736; all of these publications are incorporated by reference herein. The above ribozyme transcription units can be incorporated into a variety of vectors for introduction into mammalian cells, including but not restricted to, plasmid DNA vectors, viral DNA vectors (such as adenovirus or adeno-associated virus vectors), or viral RNA vectors (such as retroviral or alphavirus vectors) (for a review see Couture and Stinchcomb, 1996, *supra*).

30 In yet another aspect the invention features an expression vector comprising nucleic acid sequence encoding at least one of the catalytic nucleic acid molecule of the invention, in a manner which allows expression of that nucleic acid molecule. The expression vector comprises in one embodiment; a) a transcription initiation region; b) a transcription termination region; c) a gene encoding at least one said nucleic acid 35 molecule; and wherein said gene is operably linked to said initiation region and said termination region, in a manner which allows expression and/or delivery of said nucleic

acid molecule. In another preferred embodiment the expression vector comprises: a) a transcription initiation region; b) a transcription termination region; c) an open reading frame; d) a gene encoding at least one said nucleic acid molecule, wherein said gene is operably linked to the 3'-end of said open reading frame; and wherein said gene is operably linked to said initiation region, said open reading frame and said termination region, in a manner which allows expression and/or delivery of said nucleic acid molecule. In yet another embodiment the expression vector comprises: a) a transcription initiation region; b) a transcription termination region; c) an intron; d) a gene encoding at least one said nucleic acid molecule; and wherein said gene is operably linked to said initiation region, said intron and said termination region, in a manner which allows expression and/or delivery of said nucleic acid molecule. In another embodiment, the expression vector comprises: a) a transcription initiation region; b) a transcription termination region; c) an intron; d) an open reading frame; e) a gene encoding at least one said nucleic acid molecule, wherein said gene is operably linked to the 3'-end of said open reading frame; and wherein said gene is operably linked to said initiation region, said intron, said open reading frame and said termination region, in a manner which allows expression and/or delivery of said nucleic acid molecule.

#### Target Validation

One of the most challenging tasks in drug discovery is the choice of a therapeutic target. Historically, traditional biochemical and other studies have offered limited information in this regard. However, recent advances in genomics offer the potential to revolutionize both the speed and certainty of therapeutic target identification. Progress in characterizing the genes in the human genome has been very rapid, and it is now estimated that the entire complement of genes in the human genome may be sequenced before the end of this century. However, this mass of information is coming to the scientific world without a road map. Converting pure gene sequence information into a functional understanding of their role in human disease is proving to be a much more difficult problem. Even after a group of genes is associated with a particular disease, the process of validating which genes are appropriate for use as therapeutic targets is often slow and costly. Most companies with genomics activities now have access to myriad partial or full sequences, but do not possess adequate technologies to determine which of those sequences is an appropriate therapeutic target. As a result, only a few genes have been unequivocally identified as the causative agent for a specific disease.

The nucleic acid molecules of the present invention can inhibit gene expression in a highly specific manner by binding to and causing the cleavage of the mRNA

corresponding to the gene of interest, and thereby prevent production of the gene product (Christoffersen, *Nature Biotech*, 1997, 2, 483-484). Appropriate delivery vehicles can be combined with these nucleic acid molecules (including polymers, cationic lipids, liposomes and the like) and delivered to appropriate cell culture or *in vivo* animal disease 5 models as described above. By monitoring inhibition of gene expression and correlation with phenotypic results, the relative importance of the particular gene sequence to disease pathology can be established. The process may be both fast and highly selective, and allow for the process to be used at any point in the development of the organism. The 10 novel chemical composition of these nucleic acid molecules may allow for added stability and therefore increased efficacy.

#### Examples

The following are non-limiting examples demonstrating the utility of the nucleic acid molecules of the instant invention. Those in the art will recognize that certain 15 experimental conditions such as temperatures, reaction times, media conditions, transfection reagents and RNA assays are not meant to be limiting and can be readily modified without significantly altering the protocols.

#### Example 1: Identification of Potential Nucleic Acid Molecule Binding Sites

The sequences of target RNAs were screened for accessible sites using a computer 20 folding algorithm. Regions of the mRNA that did not form secondary folding structures were identified. For ribozyme sites, regions of mRNA that did not form secondary structure and contained potential hammerhead and/or hairpin cleavage sites were identified.

#### Example 2: Selection of Ribozyme Cleavage Sites in Estrogen Receptor RNA

To test whether the sites predicted by the computer-based RNA folding algorithm 25 corresponded to accessible sites in estrogen receptor. Ribozyme target sites were chosen by analyzing genomic sequences of Genbank Sequence HSERR1 (Green *et al.*, 1986, *Nature* 320, 134-139) and prioritizing the sites on the basis of folding. Ribozymes were designed that could bind each target (see Figure 3) and were individually analyzed by computer folding (Christoffersen *et al.*, 1994 *J. Mol. Struc. Theochem*, 311, 273; Jaeger *et* 30 *al.*, 1989, *Proc. Natl. Acad. Sci. USA*, 86, 7706) to assess whether the ribozyme sequences fold into the appropriate secondary structure. Those ribozymes with unfavorable intramolecular interactions between the binding arms and the catalytic core were eliminated from consideration. As noted below, varying binding arm lengths can be

chosen to optimize activity. Generally, at least 5 bases on each arm are able to bind to, or otherwise interact with, the target RNA. Hammerhead (Seq. ID. Nos. 1-1245) and hairpin ribozymes (2491-2603) are listed in tables IV and V respectively.

Example 3: Inhibition of c-raf RNA Targets using Nucleic Acid Molecules

5       Prostate cancer cells (PC-3) were grown in a growth media consisting of Kaighn's F-12K media, 10% FBS, 1% glutamine, 20 mM HEPES, and 1% pen/strep to sub-confluent densities. A 4X concentration (10  $\mu$ g/mL) of GSV (Glen Research) was prepared from a 2 mg/mL stock solution as well as a 10 $\mu$ M solution of antisense and its scrambled (mismatch) control. Complexes of antisense and GSV were formed in a 96  
10 well plate by channel pipetting in antisense and GSV to form complex solutions which are twice the final concentrations. 50  $\mu$ L of the complex solution and 50  $\mu$ L of growth medium (without antibiotics) were added to PC-3 cells and incubated for 24 hours. The final concentrations of antisense used were 400, 200, and 100 nM, while the GSV concentration was held constant at 2.5  $\mu$ g/mL. PC-3 cells were then harvested with 150  
15  $\mu$ L of RLT lysis buffer (Qiagen). RNA was purified using Qiagen's instructions and RNA was quantified using Taqman reagents and the 7700 Prism (Perkin Elmer) using the manufacturer's protocol. The c-raf RNA concentration was normalized to the c-raf RNA concentrations of the scrambled controls. The antisense sequence (Seq. I.D. No. 2717)  
20 and the data is shown in table IIIA and figure 5. The antisense molecules were capable of reducing c-raf RNA levels up to 80% compared to the mismatch control in PC-3 cells at several concentrations of antisense molecules.

Example 4: Ribozyme *in vitro* Cleavage Assay

25       Ribozymes and complementary substrates were synthesized as described above. These ribozymes can be tested for cleavage activity *in vitro*, for example using the following procedure. The ribozyme sequences are shown in figure 7.

30       *Cleavage Reactions:* Full-length or partially full-length, internally-labeled target RNA for ribozyme cleavage assay was prepared by *in vitro* transcription in the presence of [ $\alpha$ -32P] CTP, passed over a G 50 Sephadex column by spin chromatography and used as substrate RNA without further purification. Alternately, substrates may be 5'-32P-end labeled using T4 polynucleotide kinase enzyme. Assays were performed by pre-warming a 2X concentration of purified ribozyme in ribozyme cleavage buffer (50 mM Tris-HCl, pH 7.5 at 37°C, 10 mM MgCl<sub>2</sub>) and the cleavage reaction was initiated by adding the 2X ribozyme mix to an equal volume of substrate RNA (maximum of 1-5 nM) that was also pre-warmed in cleavage buffer. The assays were carried out for 1 hour at 37°C using a

final concentration of either 1  $\mu$ M ribozyme, *i.e.*, ribozyme excess. The reaction was quenched by the addition of an equal volume of 95% formamide, 20 mM EDTA, 0.05% bromophenol blue and 0.05% xylene cyanol after which the sample was heated to 95°C for 2 minutes, quick chilled and loaded onto a denaturing polyacrylamide gel. Substrate RNA and the specific RNA cleavage products generated by ribozyme cleavage were visualized on an autoradiograph of the gel. The percentage of cleavage was determined by Phosphor Imager® quantitation of bands representing the intact substrate and the cleavage products. The ribozymes were able to cleave between 1-80.8% of their complementary substrates after 2 hours (table VI).

10 Example 5: Inhibition of Cell Proliferation Using Estrogen Receptor Targeted Ribozyme

MCF-7 cells were grown in a T-75 flask to 80% confluence in growth media, which was prepared by mixing 500 mL Alpha-MEM, 10 mL 1M Hepes, 5mL sodium Pyruvate, 5 mL NEAA, 5 mL L-glutamine, 250  $\mu$ L 2.0 mg/ml insulin, 500  $\mu$ L Gentamycin, and 10% FBS following by sterile filtration.

15 Ribozyme and cationic lipid was mixed as described in example 4 with final concentrations of 20, 40, and 80 nM ribozyme and 1  $\mu$ g/mL GSV. The MCF-7 cells were treated with serum free alpha-MEM 24 hours prior to exposure to ribozyme/transfection reagent complexes. The complexes were added to the cells and allowed to continuously transfect for 1,2 , 3, 4, and 5 days. At the end of each time point, the media was removed  
20 off the cells and proliferation was measured using a CyQuant kit (Molecular Probes). Fluorescence was measured at after 10 minutes of incubation at 485 nm (excitation) and 538 (emission). Inhibition of cellular proliferation by active ribozyme was compared to inactive scrambled ribozyme controls. The sequence for the active ribozyme is given as sequence ID. No. 2725 and the inactive scrambled control is given as Seq. ID. No. 2726.  
25 The chimeric enzymatic nucleic acid was able to inhibit MCF-7 cellular proliferation at all of the tested concentrations (figure 4).

Example 6: Inhibition of Estrogen Receptor RNA Targets using antisense nucleic acid molecules

30 A 5X concentrated solution of oligonucleotide ( 1 $\mu$ M) and a 10X solution of GSV (25  $\mu$ g/mL) (Glen Research) was made prior to complexing. The 5X oligonucleotide solution (8  $\mu$ L), 10X GSV solution (40  $\mu$ L), and Optimem media (32  $\mu$ L) were mixed together and incubated at 37°C for 15 minutes. Media was aspirated off the cells followed by the addition of 80  $\mu$ L of fresh growth media (Optimem media with 10% FBS, 1% non-essential amino acids, 1% sodium pyruvate, 20mM HEPES, 1  $\mu$ g/mL insulin) and 20  $\mu$ L

of 5X complex solution. The complex was left on the cells for 20 hours and then harvested with 150  $\mu$ L of RLT lysis buffer (Qiagen). RNA levels were then quantified using Taqman reagents and the 7700 Prism (Perkin Elmer) using the manufacturer's protocol. This cellular delivery assay was used for several RPI targets in varying cell lines 5 and the data is shown in table IIIA. The levels of target RNA were either normalized to mismatch control RNA or to an internal housekeeping gene such as actin. Antisense nucleic acid molecules, which are given as Seq. ID. Nos. 2718-2723, were able to knock down estrogen receptor RNA by varying degrees. The levels of RNA inhibition ranged from 48-64% depending on the antisense sequence.

10 Example 7: Inhibition of Estrogen Receptor RNA using Ribozymes

Ribozymes and GSV transfection reagents were mixed together using a protocol similar to the one found in example 6. Target RNA was purified using a Qiagen kit. RNA levels were then quantified using Taqman reagents and the 7700 Prism (Perkin Elmer) using the manufacturer's protocol. The ribozyme specific to estrogen receptor is given as 15 sequence ID. No. 2724 and was able to inhibit the gene by 50%.

Example 8: Inhibition of c-Raf mRNA and Cellular Proliferation Using Antisense Nucleic Acid Molecules

24 Hour RNA endpoint assay: The effect of targeting intron, exon, and intron-exon junction sites with antisense molecules was tested on c-raf RNA using *in vitro* cell culture assays. Seven chimeric antisense oligonucleotides with various chemical modifications were tested and the results were compared to an untreated control and an oligonucleotide modified at every nucleotide position with a phosphorothioate modification. Prostate cancer cells (PC-3) were grown in growth media consisting of Kaighn's F-12K media, 10% FBS, 1% glutamine, 20 mM HEPES, and 1% pen/strep in 96 well plates (15,000 cells per well). For the 24 hour delivery experiments, 2.5 mg/mL of GSV transfection reagent (Glen Research) was complexed to either 200 or 400 nM concentration of antisense molecules. The complex was left on the cells for 24 hours and the cells were then harvested with 150  $\mu$ L of RLT lysis buffer (Qiagen). RNA levels were quantified using Taqman reagents and the 7700 Prism (Perkin Elmer) using the manufacturer's protocol. 25 C-raf RNA levels were normalized to actin controls and the data is shown in Figure 8. All compounds demonstrated an ability to inhibit c-raf message. Of the compounds tested, sequence I.D. 2732, 2736 and 2737 seemed particularly effective.

Sustained delivery RNA endpoint assay: Prostate cancer cells (PC-3) were grown in growth media consisting of Kaighn's F-12K media, 10% FBS, 1% glutamine, 20 mM

HEPES, and 1% pen/strep in 96 well plates (15,000 cells per well). The cells were plated to 2500 cells per well in a 96 well plate and incubated at 37°C. Antisense nucleic acid molecules (Seq. I.D. Nos. 2738, 2737, 2744; Table IV) and GSV transfection reagent were complexed as in described under example 4 to a final concentration of antisense molecules 5 of 100, 200 or 400 nM with 1.0 µg/mL of GSV transfection reagent. For seq. I.D. No. 2744, a mismatch control with similar base composition was also tested as a control. The complex was left on the cells and allowed to continuously transfect for 1, 3, or 5 days. At the end of each time period, the cells were harvested with 150 µL of RLT lysis buffer 10 (Qiagen); and RNA levels quantified using Taqman reagents and the 7700 Prism (Perkin Elmer) using the manufacturer's protocol. The inhibition of c-raf expression by molecules 15 represented by Seq. I.D. Nos. 2738, 2737, and 2744 (table IV) are shown in Figures 9, 10, and 11 respectively. All three of these molecules demonstrated the ability to reduce c-raf RNA between 60-90% compared to the untreated control.

*Proliferation Assay:* Chimeric oligonucleotides and delivery reagents were mixed 15 as described in example 6 where the final concentration of antisense oligonucleotide is 200 nM (Seq. I.D. No. 2738 and 2741) and GSV is 1 µg/mL. A mismatch antisense control was also tested (Seq. I.D. No. 2739) in this experiment. The MCF-7 cells were treated with serum free alpha-MEM 24 hours prior to exposure to antisense/transfection reagent complexes. The complexes were added to the cells and allowed to continuously transfect 20 for 1, 3, or 5 days. At the end of each time point, the media was removed off the cells and proliferation was measured using a CyQuant kit (Molecular Probes). Fluorescence was measured after 10 minutes of incubation at 485 nm (excitation) and 538 (emission). Inhibition of cellular proliferation by active oligonucleotide was compared to scrambled mismatch controls. The antisense molecules targeting c-raf mRNA were able to inhibit 25 cellular proliferation by up to 55%.

*Varying Chemical Modifications:* Referring to Figure 13 and Table IV, alterations 30 of the chemical composition of antisense molecules were made while keeping the oligonucleotide sequence constant. All of the sequences were the same except for the mismatch controls and Seq. I.D. Nos. 2738 and 2739 (table IV) which were two nucleotides shorter than the others. Antisense molecules and GSV transfection reagent were mixed using the protocol described under example 6. The complexes were added to the cells and allowed to continuously transfect for 24 hours. The cells were then harvested with 150 µL of RLT lysis buffer (Qiagen). RNA levels were then quantified using Taqman reagents and the 7700 Prism (Perkin Elmer) using the manufacturer's protocol. 35 All c-raf levels were normalized using actin RNA as a control. Referring to Figure 13,

antisense molecules which utilized phosphorothioate modifications, inverted abasic caps, 2'-O-methyl or 2'-O-methylthiomethyl modifications inhibit c-raf mRNA.

*Dose dependent Inhibition:* 0, 50, 100 150, and 200 nM of antisense molecules (seq. I.D. No. 2738 or 2737) were mixed with mismatch control antisense molecules to 5 give a final antisense/mismatch antisense concentration of 200nM for each sample. The antisense nucleic acid and GSV were complexed as in example 6 with a final GSV transfection reagent concentration at 2.5  $\mu$ g/mL. The complexes were added to the cells and allowed to continuously transfect for 24 hours. At the end of each time period, the cells were harvested with 150  $\mu$ L of RLT lysis buffer (Qiagen). RNA levels were then 10 quantified using Taqman reagents and the 7700 Prism (Perkin Elmer) using the manufacturer's protocol. All c-raf levels were normalized using actin RNA as a control. The results (Figure 14) show that c-raf is inhibited in a dose dependent manner and that the IC50 is approximately 35 nM for each of the antisense molecules.

Example 9: Ribonuclease Protection Assay of MCF-7 Cells treated with Antisense Nucleic Acid Molecules Targeting BCL-2

MCF-7 cells were plated in RPMI 1640 (10% FBS, 1% l-glutamine, 20 mM HEPES) at 100,000 cells per well for 24 hours. On the following day, the cells were treated with 150 nM of antisense nucleic acid molecules (Seq. I.D. Nos. 2749-2753; Table VIII) complexed with 5.4  $\mu$ M of LipofectAMINE (LFA) for 4 hours. The antisense/LFA 20 complex was then aspirated off and fresh RPMI 1640 media was added to the cells. 24 hours later the cells were harvested using RLT lysis buffer (Qiagen). A ribonuclease protection assay (Ambion) was then performed using the manufacturers protocol to quantitate RNA levels and then harvested with 150  $\mu$ L of RLT lysis buffer (Qiagen). RNA levels were then quantified using Taqman reagents and the 7700 Prism (Perkin 25 Elmer) using the manufacturer's protocol. Bcl-2 RNA levels were normalized to GAPDH controls and is shown in Figure 15. The antisense oligonucleotides specifically inhibited Bcl-2 expression in MCF-7 cells.

Example 10: Inhibition of k-ras in DLD-1 Cells

96-well plates with DLD-1 cells at 10,000 cells per well were plated in complete 30 RPMI 1640 media ((10% FBS, 1% l-glutamine, 20 mM HEPES, 1% pen/strep). Antisense molecules (Seq. I.D. Nos. 2754-2757; Table VIII) were complexed with GSV transfection reagent (Glen Research) using the method described in example 6. The final concentrations delivered to the cells were 200 nM antisense oligonucleotide and 1.25  $\mu$ g/mL of GSV. The complex was added to the cells for 26 hours and at the end of the

time period, the cells were harvested with 150  $\mu$ L of RLT lysis buffer (Qiagen). RNA levels were then quantified using Taqman reagents and the 7700 Prism (Perkin Elmer) using the manufacturer's protocol. All k-ras levels were normalized using actin RNA as a control. The data (Figure 16) demonstrates that the antisense molecules can inhibit 5 approximately 50-90% k-ras expression compared to the untreated or mismatch controls.

Example 11: Inhibition of Estrogen Receptor mRNA Using Antisense Molecules of varying Chemical Composition

Alteration of the chemical composition of antisense molecules were made while keeping the oligonucleotide sequence constant. All of the sequences (Table VII: Seq. I.D. 10 Nos. 2758, 2760, 2762, 2764, 2766, 2768) were the same except for the mismatch controls (Table VII: Seq. I.D. Nos. 2759, 2761, 2763, 2765, 2767, 2769). Antisense molecules and GSV transfection reagent was mixed using the protocol described in example 6. The complexes were added to MCF-7 cells and allowed to continuously transfect for 24 hours. The cells were harvested with 150  $\mu$ L of RLT lysis buffer (Qiagen). RNA levels were 15 then quantified using Taqman reagents and the 7700 Prism (Perkin Elmer) using the manufacturer's protocol. All estrogen receptor mRNA levels were normalized using actin RNA as a control. The antisense molecules which included phosphorothioate modifications, inverted abasic caps, 2'-O-methyl or 2'-O-methylthiomethyl modifications appeared to decrease estrogen receptor RNA (Figure 17).

20 Example 12: Synthesis of 3'-deoxy-3'-thio Guanosine and it's 3'-Thiophosphoramidite

Referring to Figures 18, Applicant has developed an efficient method for the synthesis of 3'-deoxy-3'-thio guanosine (13) and its 3'-thiophosphoramidite 23 from guanosine. Reaction of suitably protected guanosine with a-acetoxyisobutyryl bromide (a-AIBBr) afforded stereoselectively 3'-deoxy-3'-bromo-2'-O-acetyl-b-D-xylofuranosyl 25 derivative 3 which was converted into the 7:3 mixture of S-acyl ribofuranosyl derivatives 5 (or 6) and 3',4'-unsaturated derivative 4. S-acylated derivatives 5 and 6 were then converted in three steps into 3'-deoxy-3'-S-pyridylsulfanyl-5'-O-(4,4'-dimethoxytrityl)guanosine (11) which served as a common intermediate for the preparation of free nucleoside 13 and 3'-thiophosphoramidite 23.

30 Oligonucleotides containing 3'-S-phosphorothiolate linkage have attracted increasing interest as probes for studying the interaction of nucleic acids and their processing enzymes. In particular these analogs have been used in revealing the involvement of metal ions in phosphoester transfer reactions catalyzed by RNA (Piccirilli *et al.*, *J. Am. Chem. Soc.* 1996, 118, 10341) and ribonucleoprotein enzymes (Sontheimer

et al., *Nature* 1997, 308, 801). The synthesis of 3'-S-phosphorothiolate linked deoxyribodinucleotides have been reported using solution chemistry and solid phase chemistry (Cosstick et al., *Nucleic Acids. Res.* 1990, 18, 829; Li et al., *Tetrahedron* 1992, 48, 2729; Li et al., *J. Chem. Soc. Perkin Trans. 1* 1994, 2123; Vyle et al., *Biochemistry* 1992, 31, 3012).

5 The synthesis of ribonucleotide 3'-S-phosphorothiolate analogs has been limited to the preparation of UspU (Liu et al., *Tetrahedron Lett.* 1996, 37, 925) and IspU (Weinstein et al., *J. Am. Chem. Soc.* 1996, 118, 10341) dimers using solution chemistry. Recently, Sun et al. (*RNA* 1997, 3, 1352) described direct incorporation of 3'-S-10 phosphorothioamidites into RNA using standard phosphoramidite solid phase synthesis.

One general approach to the synthesis of 3'-thio ribonucleosides involves preparation of 3-thio ribose derivative, followed by the attachment of the desired nucleoside base (Ryan et al., *J. Org. Chem.* 1968, 33, 1783; Cao, et al., *Bioorg. Med. Chem. Lett.* 1994, 4, 807). While glycosylation reactions using pyrimidines proceed in 15 high yields, purine bases generally give more complex mixtures because both N-7 and N-9 of the purine base are reactive towards glycosylation. Sun et al., *supra*, reported the first synthesis of 3'-thio guanosine derivatives using the above described approach. Coupling of per-acylated 3'-thiouridine with persilylated *N*<sup>2</sup>-acetylguanine proceeded in ca 40% yield and subsequent synthetic steps proceeded in a low overall yield.

20 Synthesis of 3'-thio adenosine (Mengel, et al., *Tetrahedron Lett.* 1977, 1177), 3'-thio uridine (Liu et al., 1996 *supra*) and 3'-thio inosine (Higson et al., *Tetrahedron* 1996, 52, 1027) all starting from preformed nucleosides are also reported.

Applicant describes a novel and improved process for the synthesis of 3'-deoxy-3'-thio guanosine (13) and its phosphoramidite 23 from guanosine as a starting material. It 25 was recently reported (He et al., *Tetrahedron Lett.* 1995, 39, 6991) that the reaction of *N*<sup>2</sup>-(dimethylaminomethylene)-guanosine 1 with a-AIBBr (the Mattocks-Moffatt reagent; Russell et al., *J. Am. Chem. Soc.* 1973, 95, 4025) proceeded stereoselectively yielding exclusively 3'-bromo-3'-deoxy-*b*-D-xylofuranosyl derivative. In general, reactions of base-unprotected purine nucleosides with this reagent result in the mixtures of trans bromo 30 acetates of xylo- and arabino- configuration. Applicant used this reaction on the suitably 5'-protected *N*<sup>2</sup>-(dimethylaminomethylene)guanosine derivative 2 (Scheme 1; Figure 18). 5'- protection in 2 helps to reduce the complexity of reaction products by eliminating possible formation of the mixture of 5'-OH, 5'-(2,5,5-trimethyl-1,3-dioxolan-4-on-yl) and/or 5'-*O*-acylated derivatives in reaction with a-AIBBr. This way, identification of the 35 reaction products becomes straightforward. Applicant has chosen the *t*-butyldiphenylsilyl (TBDPS) protection because of its relatively high stability towards acidic conditions

required during the reaction with a-AIBBr in moist acetonitrile. This group is also expected to undergo selective cleavage in the presence of S-acyl groups. Reaction of **1** with TBDPS-Cl proceeded quantitatively to afford the 5'-O-silyl derivative **2** which reacted smoothly with a-AIBBr yielding the desired 3'-bromo-3'-deoxy-*b*-D-xylofuranosyl derivative **3** in a high yield. Reaction of **3** with potassium thioacetate or potassium thiobenzoate yielded the 3'-S-Ac or 3'-S-Bz derivatives **5** and **6**, respectively, along with 3',4'-unsaturated derivative **4**. The latter is formed by competing elimination reaction. The ratio was 7:3 in favor of the substitution products **5** and **6** which could not be separated from the elimination product **4** at this stage. The mixture of **4** and **5** (or **4** and **6**) 10 was treated with tetrabutylammonium fluoride (TBAF) buffered with an excess of acetic acid; following chromatographic separation the desired 5' de-protected derivative **7** was obtained in a good yield. The unsaturated derivative **4** was unstable under the acidic reaction conditions and could not be isolated in a pure state by silica gel column chromatography. When triethylamine trihydrofluoride (TEA•3HF) reagent was used for 15 deprotection, desilylation did not proceed to completion.

It is desirable to keep guanosine derivatives protected with lipophylic groups during synthetic transformations because of the solubility problems encountered with unprotected derivatives. For that reason **7** and **8** were re-protected in high yields with 4,4'-dimethoxytrityl (DMT) group to give the fully protected derivatives **9** and **10**, 20 respectively. DMT group provided a hydrophobic tag which simplified work-up and purification of subsequent synthetic intermediates. Next, **9** and **10** were converted into S-(pyridyl-2-disulfanyl) derivative **11** using aqueous methylamine followed by the disulfide exchange reaction with 2,2'-dipyridyl disulfide. It is reported that the removal of 2'-O-acyl protection in ribofuranosyl derivatives similar to **9** and **10** proceeds with difficulty. 25 Applicant found that 40% aqueous methylamine easily removed all acyl protecting groups from **9** and **10** and was the base of choice because, contrary to aqueous ammonia, it completely solubilised the fully protected substrates. *In situ* protection of SH as S-pyridyl-2-disulfanyl derivative was achieved using 2,2'-dipyridyl disulfide in DMF to afford **11** in 85% yield for these two steps.

30 In order to synthesize the free nucleoside, **13**, **11** was treated with dithiothreitol (DTT) in chloroform. When triethylamine was added to the reaction mixture the reaction was faster than in it's absence but at the same time **12** was converted into its S-TEA salt. Final deprotection of the DMT group of **12** was achieved with 1N HCl in methanol in the presence of DTT which quenched the released DMT-cation. In the absence of DTT, 35 quantitative S-alkylation took place. Applicant is the first to report on the synthesis of the guanosine 3'-thio analog **13**. When unbuffered TBAF in THF was used to desilylate the

mixture of 4 and 5, *S*-Ac protecting group was also removed leading to formation of the disulfide 14 (Scheme 2; Figure 19). Under these conditions the 3',4'-unsaturated derivative 15 remained intact and was separated from 14 by silica gel column chromatography. Products were invariably contaminated with TBAF. Attempted 5 rechromatography of 15 led to its decomposition, though. Disulfide 14 was 5'-DMT protected to afford 16.

The selective removal of the 3'-acetyl group of 16 (Scheme 2; Figure 19), followed by the introduction of 2'-*O*-*t*-butyldimethylsilyl (TBDMS) protection, then reduction of 3'-disulfide and 3'-phosphitylation would be the shortest way to prepare the 10 desired 3'-thiophosphoramidite building block. Reactions of 16 with mild deacylating agents like basic ion exchangers in OH<sup>-</sup> or CN<sup>-</sup> form selectively removed 2'-*O*-acetyl protection, but at the same time nucleoside was strongly absorbed on the resin, resulting in low recoveries. Applicant used basic treatment followed by *S*-protection with *S*-pyridyl group, as used for the preparation of the *S*-pyridyl-2-disulfanyl derivative 11 from *S*-15 acylated 9 and 10. In this manner 11 was obtained from the disulfide 16 in 67% yield.

The phosphoramidite synthesis is shown in Scheme 3 (Figure 20). Reaction of 11 with *N,N*-dimethylformamide dimethyl acetal yielded the desired *N*-protected derivative 18 in 23% yield. Unfortunately, this reagent also effected the cleavage of the *S*-pyridyl protection leading to formation of disulfide 17 in 33% yield. 18 was smoothly 2' 20 protected with TBDMS group using *t*-butyldimethylsilyl trifluoromethanesulfonate (TBDMS-Tf). Alternatively, 11 was silylated with TBDMS-Cl to afford 20 and then *N*-protected using isobutyric anhydride (*i*-Bu<sub>2</sub>O) in the presence of 4-dimethylaminopyridine (DMAP) yielding the fully protected 21. In the absence of DMAP only starting material 25 was recovered. On the other hand, reaction of 20 with isobutyryl chloride led to *N*-bis-acylation. Reduction of 21 with DTT afforded 3'-SH derivative 22, which appeared as a mixture of two rotamers in <sup>1</sup>H NMR. Resonances of the major rotamer were in accordance with the ones reported by Sun *et al.* Phosphitylation of 22 under standard conditions afforded 3'-thiophosphoramidite 23. Applicant has described an efficient 30 synthesis of 3'-deoxy-3'-thio guanosine. Keeping all synthetic intermediates protected with lipophylic groups enabled their chromatographic purification and, consequently, a good recovery of the products.

#### Experimental Section

**General.** All reactions were carried out under a positive pressure of argon in anhydrous solvents. Commercially available reagents and anhydrous solvents were used 35 without further purification. <sup>1</sup>H (400.075 MHz) and <sup>31</sup>P (161.947 MHz) NMR spectra

were recorded in  $\text{CDCl}_3$ , unless stated otherwise, and chemical shifts in ppm refer to TMS and  $\text{H}_3\text{PO}_4$ , respectively. Analytical thin-layer chromatography (TLC) was performed with Merck Art.5554 Kieselgel 60 F<sub>254</sub> plates and flash column chromatography using Merck 0.040-0.063 mm silica gel 60. Mass spectra were obtained by fast atom 5 bombardment method.

**5'-O-t-Butyldiphenylsilyl-N<sup>2</sup>-(dimethylaminomethylene)guanosine (2).** To a stirred solution of  $N^2$ -(dimethylaminomethylene)guanosine (1) (5.5 g, 16.3 mmol) in pyridine (100 mL) *t*-butyldiphenylsilyl chloride (6.2 mL, 23.8 mmol) was added under argon. The reaction mixture was stirred at rt for 16 h, then quenched with methanol (20 mL) and evaporated to a syrup *in vacuo*. The residue was crystallized from ethanol-ether (9 g, 96%), mp, <sup>1</sup>H NMR (DMSO-d<sub>6</sub> + D<sub>2</sub>O) δ 8.46 (s, 1H, CH=N), 7.89 (s, 1H, H-8), 7.58-7.31 (m, 10H, Ph), 5.81 (d,  $J_{1',2'}=4.8$ , 1H, H-1'), 4.46 (app t,  $J_{2',1'}=4.8$ , 1H, H-2'), 4.23 (app t,  $J_{3',2'}=5.0$ , 1H, H-3'), 3.97 (m, 1H, H-4'), 3.84 (dd,  $J_{5',4'}=2.8$ ,  $J_{5',5'}=12.0$ , 1H, H-5'), 3.74 (dd,  $J_{5',4'}=4.4$ ,  $J_{5',5'}=12.0$ , 1H, H-5"), 3.05 (s, 3H, Me), 2.97 (s, 3H, Me), 0.94 (s, 9H, *t*-Bu), HRMS (FAB<sup>+</sup>) calcd for  $\text{C}_{29}\text{H}_{36}\text{N}_6\text{O}_5\text{Si}$  (MH<sup>+</sup>): calc 577.2595, found 577.26095.

**1-(2-O-Acetyl-5-O-t-Butyldiphenylsilyl-3-deoxy-3-bromo-*b*-D-xylofuranosyl)- $N^2$ -(dimethylaminomethylene)guanosine (3).** To a cooled (0 °C) solution of 2 (5.8 g, 10 mmol) and water (0.12 mL) in acetonitrile (130 mL) a-acetoxyisobutryl bromide (5.56 mL, 38 mmol) was added and the mixture was stirred at rt for 3 h. The solution was poured into saturated aq.  $\text{NaHCO}_3$  (100 mL) and extracted with  $\text{CH}_2\text{Cl}_2$  (3 x 200 mL). The combined organic layers were dried ( $\text{Na}_2\text{SO}_4$ ) and concentrated to give chromatographically pure white foam (6 g, 87%), <sup>1</sup>H NMR δ 8.97 (br s, 1H, NH), 8.62 (s, 1H, CH=N), 7.82 (s, 1H, H-8), 7.73-7.31 (m, 10H, Ph), 6.09 (s, 1H, H-2'), 5.92 (d,  $J_{1',2'}=1.6$ , 1H, H-1'), 4.42 (m, 1H, H-4'), 4.36 (m, 1H, H-3'), 4.06 (dd,  $J_{5',4'}=5.6$ ,  $J_{5',5'}=10.4$ , 1H, H-5'), 3.97 (dd,  $J_{5',4'}=6.4$ ,  $J_{5',5'}=10.4$ , 1H, H-5"), 3.17 (s, 3H, *N*-Me), 3.07 (s, 3H, *N*-Me), 2.19 (s, 3H, *O*-Ac), 1.07 (s, 9H, *t*-Bu), HRMS (FAB<sup>+</sup>) calcd for  $\text{C}_{31}\text{H}_{37}\text{BrN}_6\text{O}_5\text{Si}$  (MH<sup>+</sup>): calc 681.1856, found 681.1850.

**1-(2-O-Acetyl-5-O-t-butyldiphenylsilyl-3-deoxy-*b*-D-glycero-pent-3-enofuranosyl- $N^2$ -(dimethylaminomethylene)guanosine (4) and 2'-O-acetyl-5'-O-t-butyldiphenylsilyl-3'-deoxy-3'-S-thioacetyl- $N^2$ (dimethylaminomethylene)- guanosine (5).** 3 (5.4 g, 7.9 mmol) was dissolved in dry DMF (50 mL) and potassium thioacetate (2.7 g, 23.6 mmol) was added to the solution. The reaction mixture was stirred at 60 °C for 16 h and then evaporated to a syrup under reduced pressure. The residue was partitioned between aq.  $\text{NaHCO}_3$ -brine 1:1 solution and dichloromethane, the organic layer was dried ( $\text{Na}_2\text{SO}_4$ ), evaporated to dryness and chromatographed on the column of silicagel using 2-

10% gradient of methanol in dichloromethane **4** and **5** co-eluted yielding, after evaporation, a yellowish foam (4.8 g). <sup>1</sup>H NMR indicated a 3:7 ratio of **4** to **5**.

When potassium thiobenzoate was used instead of potassium thioacetate an inseparable mixture of **4** and **2'-O-acetyl-5'-O-t-butylidiphenylsilyl-3'-deoxy-3'-S-thiobenzoyl-N<sup>2</sup>(dimethylaminomethylene)guanosine (6)** was obtained in a similar yield and ratio to the unsaturated derivative **4** as above.

**2'-O-Acetyl-3'-deoxy-3'-S-thioacetyl-N<sup>2</sup>(dimethylaminomethylene) guanosine (7).** The above mixture of **4** an **5** (0.9 g) was dissolved in THF (15 ml) and acetic acid (0.37 ml, 6.5 mmol) was added followed by TBAF•3H<sub>2</sub>O (0.82 g, 2.6 mmol). The reaction mixture was stirred at rt for 5 h, then diluted with dichloromethane, washed with water and 10% aq. NaHCO<sub>3</sub>. Aqueous layers were back-washed with dichloromethane, organic layers were combined, dried (Na<sub>2</sub>SO<sub>4</sub>) and evaporated to dryness. Silica gel column chromatography using 2-10% gradient of methanol in dichloromethane afforded **7** as a yellowish foam (300 mg, ca 74%), <sup>1</sup>H NMR d 8.87 (br s, 1H, NH), 8.78 (s, 1H, CH=N), 7.73 (s, 1H, H-8), 5.80 (d, J<sub>1',2'</sub>=2.0, 1H, H-1'), 5.76 (dd, J<sub>2',1'</sub>=2.0, J<sub>2',3'</sub>=6.4, 1H, H-2'), 4.94 (dd, J<sub>3',2'</sub>=6.4, J<sub>3',4'</sub>=9.6, 1H, H-3'), 4.22 (d, J<sub>4',3'</sub>=9.6, 1H, H-4'), 4.04 (br s, 1H, 5'-OH), 3.99 (d, J<sub>5',5''</sub>=12.0, 1H, H-5'), 3.71 (d, J<sub>5'',5'</sub>=12.0, 1H, H-5''), 3.19 (s, 3H, N-Me), 3.04 (s, 3H, N-Me), 2.34 (s, 3H, S-Ac), 2.13 (s, 3H, O-Ac), HRMS (FAB<sup>+</sup>) calcd for C<sub>17</sub>H<sub>22</sub>N<sub>6</sub>O<sub>6</sub>S (MH<sup>+</sup>): 439.1355, found 439.1405.

**2'-O-Acetyl-3'-deoxy-3'-S-thiobenzoyl-N<sup>2</sup>(dimethylaminomethylene) guanosine (8).** Using the same procedure as above, **8** was synthesized from the mixture of **4** and **6** in ca 70% yield, <sup>1</sup>H NMR d 8.90 (br s, 1H, NH), 8.57 (brs, 1H, NH), 7.69 (s, 1H, H-8), 5.88 (m, 2H, H-1',H-2'), 5.30 (m, 1H, H-3'), 4.34 (d, J<sub>4',3'</sub>=9.2, 1H, H-4'), 4.04 (d, J<sub>5',5''</sub>=12.8, 1H, H5'), 3.80 (dd, J<sub>5'',OH</sub>=9.6, J<sub>5'',5'</sub>=12.8, 1H, H-5''), 3.69 (br s, 1H, 5'-OH), 3.25 (s, 3H, N-Me), 3.10 (s, 3H, N-Me), 2.16 (s, 3H, O-Ac), HRMS (FAB<sup>+</sup>) calcd for C<sub>17</sub>H<sub>22</sub>N<sub>6</sub>O<sub>6</sub>S (MH<sup>+</sup>): 501.1556, found 501.1561.

**2'-O-Acetyl-3'-deoxy-3'-S-thioacetyl-5'-O-(4,4'-dimethoxytrityl)-N<sup>2</sup>(dimethylamino-methylene)guanosine (9).** **7** (720 mg, 1.64 mmol) was dissolved in dry pyridine (15 ml) and DMT-Cl (1.1 g, 3.3 mmol) was added. The reaction mixture was stirred at rt for 4 h, quenched with methanol and evaporated to a syrup which was partitioned between 5% aq. NaHCO<sub>3</sub> and CH<sub>2</sub>Cl<sub>2</sub>. Organic layer was washed with brine, dried (Na<sub>2</sub>SO<sub>4</sub>) and evaporated to dryness *in vacuo*. The residue was purified by silica gel column chromatography using 1-5% gradient of methanol in dichloromethane to yield the product as a colorless foam (0.85 g, 70%), <sup>1</sup>H NMR d 8.69 (s, 1H, CH=N), 8.58 (br s, 1H, NH), 7.69 (s, 1H, H-8), 7.38-6.74 (m, 13H, H-8, aromatic), 6.06 (dd, J<sub>2',3'</sub>=6.4, J<sub>2',1'</sub>=1.2, 1H, H-2'), 5.82 (d, J<sub>1',2'</sub>=1.2, 1H, H-1'), 4.73 (dd, J<sub>3',4'</sub>=10.6, J<sub>3',2'</sub>=6.4, 1H, H-3'), 4.21

(dq,  $J_{4',3}=10.6$ ,  $J_{4',5}=3.0$ ,  $J_{4',5''}=4.4$ , 1H, H-4'), 3.78 (s, 6H, 2xOMe), 3.36 (m, 2H, H-5', H-5''), 3.07 (s, 3H, N-Me), 3.05 (s, 3H, N-Me), 2.26 (s, 3H, S-Ac), 2.15 (s, 3H, O-Ac), HRMS (FAB<sup>+</sup>) calcd for  $C_{38}H_{40}N_6O_8S$  ( $MH^+$ ): 741.2707, found 741.2692.

**2'-O-Acetyl-3'-deoxy-3'-S-thiobenzoyl-5'-O-(4,4'-dimethoxytrityl)-**

5  **$N^2$ (dimethyl-aminomethylene)guanosine (10).** Using similar procedure as described above, **8** was converted into **10** in 69% yield, <sup>1</sup>H NMR d 8.80 (s, 1H, CH=N), 8.65 (br s, 1H, NH), 7.70 (s, 1H, H-8), 7.88-6.66 (m, 19H, aromatic), 6.17 (d,  $J_{2',3}=5.8$ , 1H, H-2'), 5.86 (d,  $J_{1',2}=1.2$ , 1H, H-1'), 5.08 (dd,  $J_{3',4}=10.4$ ,  $J_{3',2}=5.8$ , 1H, H-3'), 4.31 (m, 1H, H-4'), 3.67 (s, 6H, 2xOMe), 3.45 (m, 2H, H-5', H-5''), 3.06 (s, 6H, 2XN-Me) 2.15 (s, 3H, O-Ac), HRMS (FAB<sup>+</sup>) calcd for  $C_{43}H_{42}N_6O_8S$  ( $MH^+$ ): 803.2863, found 803.2855.

10 **3'-Deoxy-3'-S-pyridylsulfanyl-5'-O-(4,4'-dimethoxytrityl)-guanosine (11).** A. **9** (530 mg, 0.38 mmol) was dissolved in 40% aqueous methylamine (50 ml) and the mixture is kept at rt for 16 h. The solvent is removed in *vacuo* and the residual syrup dissolved in argon purged DMF (30 ml) containing 2,2'-dipyridyl disulfide (340 mg, 1.54 mmol). The 15 reaction mixture is heated at 60 °C for 10 h and then evaporated to a syrup *in vacuo*. Column chromatography on silica gel using 1-12% gradient of methanol in dichloromethane afforded **11** as a colorless solid (460 mg, 85%), <sup>1</sup>H NMR d 10.64 (br s, 1H, NH), 8.39 (m, 1H, Pyr), 7.83 (s, 1H, H-8), 7.73-6.72 (m, 16H, aromatic), 6.50 (d,  $J_{OH,2}=4.80$ , 1H, OH-2'), 6.45 (br s, 2H, NH<sub>2</sub>), 5.81 (d,  $J_{1',2}=2.4$ , 1H, H-1'), 4.83 (m, 1H, H-2'), 4.34 (m, 1H, H-4'), 4.09 (dd,  $J_{3',2}=6.00$ ,  $J_{3',4}=7.8$ , 1H, H-3'), 3.70 (s, 6H, 2XOMe), 3.11 (dd,  $J_{5',5}=11.2$ ,  $J_{5',4}=4.8$ , 1H, H-5''), HRMS (FAB<sup>+</sup>) calcd for  $C_{36}H_{34}N_6O_6S_2$  ( $MH^+$ ): 711.2060, found 711.2076.

B. Using the same procedure as above, but starting from *S*-benzoyl derivative **10**, **11** was prepared in 80% yield.

25 C. Starting from **16** (830 mg, 1.12 mmol) and using the above conditions **11** (570 mg, 67%) was obtained.

30 **3'-Deoxy-3'-thio-5'-O-(4,4'-dimethoxytrityl)-guanosine (12).** To the solution of **11** (240 mg, 0.34 mmol) in chloroform (14 ml) dithiothreitol (DTT) (125 mg, 0.81 mmol) was added and the reaction mixture was stirred at rt for 3 h. It was then evaporated to a 35 syrup *in vacuo*, and the product was precipitated by addition of peroxide-free ether, precipitate was filtered off, washed with ether and dried (230 mg of the crude material), <sup>1</sup>H NMR (DMSO-d<sub>6</sub>) d 10.63 (br s, 1H, NH), 7.86 (s, 1H, H-8), 7.32-6.80 (m, 13H, aromatic), 6.49 (br s, 2H, NH<sub>2</sub>), 5.81 (s, 1H, H-1'), 4.43 (d,  $J_{2',3}=4.8$ , 1H, H-2'), 3.93 (m, 1H, H-4'), 3.79 (dd,  $J_{3',2}=4.8$ ,  $J_{3',4}=9.6$ , 1H, H-3'), 3.71 (s, 6H, 2XOMe), 3.16 (dd,  $J_{5',5}=10.4$ ,  $J_{5',4}=4.8$ , 1H, H-5'').

**3'-Deoxy-3'-thio-guanosine (13).** The mixture of the crude **12** (230 mg, 0.33 mmol) and DTT (150 mg) was dissolved in 1 N methanolic HCl (12 ml) and the reaction mixture was kept at rt for 3 h. It was then concentrated *in vacuo* and the residue coevaporated with toluene two times. Addition of ethyl acetate afforded precipitate which 5 was filtered off, washed well with ethyl acetate and dried to afford **13** (90 mg, 79%). The product was reprecipitated from water, <sup>1</sup>H NMR (CD<sub>3</sub>OD) δ 8.10 (s, 1H, H-8), 5.91 (s, 1H, H-1'), 4.37 (d, J<sub>2',3'</sub>=5.2, 1H, H-2'), 3.97 (m, 2H, H-4', H-5'), 3.82 (dd, J<sub>5'',5'</sub>=13.0, J<sub>5'',4'</sub>=3.4, 1H, H-5''), 3.64 (dd, J<sub>3',2'</sub>=5.2, J<sub>3',4'</sub>=9.6, 1H, H-3'), HRMS (FAB<sup>+</sup>) calcd for C<sub>10</sub>H<sub>13</sub>N<sub>5</sub>O<sub>4</sub>S (MH<sup>+</sup>): 300.0767, found 300.0767.

10 **Bis (2-O-acetyl-N<sup>2</sup>-(dimethylaminomethylene)guanosin-3-yl)disulfide (14 ) and 1-(2-O-acetyl-3-deoxy-**b**-D-glycero-pent-3-enofuranosyl)-N<sup>2</sup>-(dimethylaminomethylene) guanosine (15 ).** The mixture of **4** and **5** (4.8 g) was dissolved in THF (100 mL) and 1M TBAF in THF (10 mL) was added. The reaction mixture was stirred for 3 h at rt and then evaporated to a syrup *in vacuo*. Silica gel column chromatography using 15 2-10% gradient of methanol in dichloromethane yielded the faster eluting **15** (1 g, 35% for two steps from **3**, colorless foam), <sup>1</sup>H NMR δ 8.96 (br s, 1H, NH), 8.57 (s, 1H, CH=N), 7.65 (s, 1H, H-8), 6.41 (s, 1H, H-2'), 6.04 (s, 1H, H-1'), 5.42 (m, 1H, H-3'), 4.32 (m, 2H, H-5',H-5''), 3.19 (s, 3H, N-Me), 3.06 (s, 3H, N-Me), 2.11 (s, 3H, Ac). The slower eluting 14 was obtained as a yellowish solid (0.9 g, 29% for two steps), <sup>1</sup>H NMR (DMSO-d<sub>6</sub>) δ 20 11.34 (br s, 1H, NH), 8.53 (s, 1H, CH=N), 8.00 (s, 1H, H-8), 5.97 (d, J<sub>1',2'</sub>=2.4, 1H, H-1'), 5.89 (dd, J<sub>2',1'</sub>=2.4, J<sub>2',3'</sub>=6.0, 1H, H-2'), 5.23 (t, J<sub>OH,S'</sub>=5.6, 1H, 5'-OH), 4.11 (m, 1H, H-4'), 4.02 (dd, J<sub>3',2'</sub>=6.0, J<sub>3',4'</sub>=8.4, 1H, H-3'), 3.78 (dm, J<sub>5',5''</sub>=12.0, 1H, H-5'), 3.60 (dm, J<sub>5'',5'</sub>=12.0, 1H, H-5''), 3.10 (s, 3H, N-Me), 3.00 (s, 3H, N-Me), 2.06 (s, 3H, O-Ac), HRMS (FAB<sup>+</sup>) calcd for C<sub>30</sub>H<sub>38</sub>N<sub>12</sub>O<sub>10</sub>S<sub>2</sub> (MH<sup>+</sup>): 791.2354, found 791.2355.

25 **Bis (2-O-acetyl-5-O-(4,4'dimethoxytrityl)-N<sup>2</sup>-(dimethylaminomethylene) guanosin-3-yl) disulfide (16).** To the solution of **14** (400 mg, 0.5 mmol) in dry pyridine (10 ml) DMT-Cl (508 mg, 1.5 mmol) was added and the mixture was stirred 4 h at rt. Methanol (10 ml) was added and the solution evaporated to dryness. The residue is partitioned between saturated NaHCO<sub>3</sub> and dichloromethane, organic layer washed with 30 brine, dried (Na<sub>2</sub>SO<sub>4</sub>) and evaporated to a syrup. Silica gel column chromatography using 2-10% gradient of methanol in dichloromethane yielded product as a yellowish foam (620 mg, 71% yield), <sup>1</sup>H NMR δ 8.72 (br s, 1H, NH), 8.01 (s, 1H, CH=N), 7.48-7.21 (m, 14H, H-8, aromatic), 6.25 (d, J<sub>2',3'</sub>=4.8, 1H, H-2'), 5.78 (s, 1H, H-1'), 4.00 (m, 2H, H-3', H-4'), 3.78 (s, 6H, 2xOMe), 3.50 (br s, 2H, H-5',H-5''), 3.14 (s, 3H, N-Me), 3.13 (s, 3H, N-Me), 35 1.82 (s, 3H, O-Ac), HRMS (FAB<sup>+</sup>) calcd for C<sub>72</sub>H<sub>74</sub>N<sub>12</sub>O<sub>14</sub>S<sub>2</sub> (MH<sup>+</sup>): 1395.4967, found 1395.4943.

**Bis (5-*O*-(4,4'dimethoxytrityl)-*N*<sup>2</sup>-(dimethylaminomethylene) guanosin-3-yl)-disulfide (17).** A. 16 (60 mg, 0.04 mmol) is dissolved in dry methanol and ion exchange resin AG 1X8 (OH<sup>-</sup>) (1 g) is added. The mixture was stirred at 55 °C for 16 h, the resin was filtered off and washed well with hot methanol. The filtrate was evaporated to dryness *in vacuo* yielding pure 17 as a colorless solid (16 mg, 28%), <sup>1</sup>H NMR (DMSO-d<sub>6</sub>) d 11.34 (br s, 1H, NH), 8.48 (s, 1H, CH=N), 7.92 (s, 1H, H-8), 7.31-6.73 (m, 13H, aromatic), 6.27 (d, J<sub>OH,2</sub>=5.2, 1H, 2'-OH), 5.88 (d, J<sub>1',2</sub>=1.2, 1H, H-1'), 4.60 (m, 1H, H-2'), 4.16 (m, 1H, H-4'), 4.08 (m, 1H, H-3'), 3.65 (s, 6H, 2XOMe), 3.65 (m, 2H, H-5', H-5''), 3.02 (s, 3H, N-Me), 2.97 (s, 3H, N-Me), HRMS (FAB<sup>+</sup>) calcd for C<sub>68</sub>H<sub>70</sub>N<sub>12</sub>O<sub>12</sub>S<sub>2</sub> (MH<sup>+</sup>): 1311.4756, found 1311.4746.

B. Using Amberlyst A-26 (CN<sup>-</sup>) under the the above reaction conditions, 17 was obtained from 16 in 21% yield.

**3'-Deoxy-3'-*S*-pyridylsulfanyl-5'-*O*-(4,4'-dimethoxytrityl)-*N*<sup>2</sup>-(dimethylaminomethylene)guanosine (18) and 17.** To the solution of 11 (400 mg, 0.56 mmol) in dry pyridine (5 ml), *N,N*-dimethylformamide dimethyl acetal (1.2 ml, 9 mmol) was added and the reaction mixture was stirred at rt for 16 h. Solvents were removed in *vacuo* and the residue chromatographed on the column of silica gel using 1-50% gradient of methanol in dichloromethane. Fractions containing the faster running material were combined and concentrated *in vacuo* to give 18 (110 mg, 23%), <sup>1</sup>H NMR d 8.73 (br s, 1H, NH), 8.53 (s, 1H, CH=N), 8.49 (m, 1H, pyridine), 7.71 (s, 1H, H-8), 7.62 (m, 1H, pyridine), 7.44-6.79 (m, 15H, aromatic), 6.09 (s, 1H, H-1'), 4.55 (d, J<sub>2',3</sub>=4.8, 1H, H-2'), 4.23 (dq, J<sub>4',3</sub>=10.5, J<sub>4',5</sub>=2.8, J<sub>4',5''</sub>=3.4, 1H, H-4'), 4.14 (dd, J<sub>3',2</sub>=4.8, J<sub>3',4</sub>=10.5, 1H, H-3'), 3.78 (s, 6H, 2XOMe), 3.57 (dd, J<sub>5',5''</sub>=10.6, J<sub>5',4</sub>=2.8, 1H, H-5'), 3.41 (dd, J<sub>5',5''</sub>=10.6, J<sub>5'',4</sub>=3.4, 1H, H-5''), 3.08 (s, 3H, N-CH<sub>3</sub>), 3.05 (s, 3H, N-CH<sub>3</sub>). Fractions containing the slower running compound were collected and evaporated to dryness *in vacuo* to give 17 as a colorless foam (120 mg, 33%), <sup>1</sup>H NMR identical to that of 17 obtained using the above procedures.

**2'-*O*-*t*-Butyldimethylsilyl-3'-deoxy-3'-*S*-pyridylsulfanyl-5'-*O*-(4,4'-dimethoxytrityl)-*N*<sup>2</sup>-(dimethylamino-methylene) guanosine (19).** 18 (110 mg, 0.14 mmol) was dissolved in dry pyridine (1 ml) and TBDMS-Tf (0.103 ml, 0.45 mmol) was added to the solution. The reaction mixture was stirred at rt for 5 h, then quenched with methanol and evaporated to a syrup *in vacuo*. The residue was dissolved in dichloromethane, washed with 5% aqueous NaHCO<sub>3</sub>, then brine and the organic layer was dried (Na<sub>2</sub>SO<sub>4</sub>) and concentrated to the syrup. Column chromatography on silica gel using 1-10% gradient of methanol in ethyl acetate afforded 19 as a colorless solid (90 mg, 71%), <sup>1</sup>H NMR d 8.69 (br s, 1H, NH), 8.50 (s, 1H, CH=N), 8.37 (m, 1H, pyridine), 7.81

(s, 1H, H-8), 7.50-6.74 (m, 16H, aromatic), 5.98 (d,  $J_{1',2'}=2.4$ , 1H, H-1'), 4.75 (dd,  $J_{2',3'}=5.0$ ,  $J_{2',1'}=2.4$ , 1H, H-2'), 4.50 (m, 1H, H-4'), 3.99 (dd,  $J_{3',2'}=5.0$ ,  $J_{3',4'}=8.4$ , 1H, H-3'), 3.76 (s, 6H,  $2\text{XOCH}_3$ ), 3.62 (dd,  $J_{5',5''}=10.9$ ,  $J_{5',4'}=2.2$ , 1H, H-5'), 3.38 (dd,  $J_{5',5''}=10.9$ ,  $J_{5',4'}=4.4$ , 1H, H-5''), 3.06 (s, 3H,  $N\text{CH}_3$ ), 3.04 (s, 3H,  $N\text{CH}_3$ ), 0.93 (s, 9H, *t*-Bu), 0.17 (s, 5H, Me), 0.10 (s, 3H, Me), HRMS (FAB<sup>+</sup>) calcd for  $\text{C}_{45}\text{H}_{53}\text{N}_7\text{O}_6\text{S}_2\text{Si} (\text{MH}^+)$  880.3346, found 880.3357.

**2'-*t*-Butyldimethylsilyl-3'-deoxy-3'-S-pyridylsulfanyl-5'-O-(4,4'-dimethoxytrityl) guanosine (20).** 3'-Deoxy-3'-S-pyridylsulfanyl-5'-O-(4,4'-dimethoxytrityl)guanosine 11 (410 mg, 0.58 mmol) was dissolved in dry pyridine (36 ml) and imidazole (2.36 g, 35 mmol) and TBDMS-Cl (4.29 g, 28 mmol) were added. The reaction was stirred at rt 16 h, then evaporated to a syrup in *vacuo*. The residue was partitioned between dichloromethane and saturated aq.  $\text{NaHCO}_3$ , organic layer washed with water, dried ( $\text{Na}_2\text{SO}_4$ ) and concentrated to a syrup. Column chromatography on silica gel using 1-10% gradient of methanol in dichloromethane afforded the product as a white foam (430 mg, 85%), <sup>1</sup>H NMR (DMSO-d<sub>6</sub>) δ 11.99 (br s, 1H, NH), 8.39 (m, 1H, Pyr), 7.84 (s, 1H, H-8), 7.71 (m, 1H, Pyr), 7.62-6.72 (m, 15H, aromatic) 6.41 (br s, 2H,  $\text{NH}_2$ ), 5.80 (d,  $J_{1',2'}=4.4$ , 1H, H-1'), 5.04 (app t,  $J_{2',3'}=4.4$ , 1H, H-2'), 4.39 (m, 1H, H-4'), 4.01 (app t,  $J_{3',4'}=6.4$ , 1H, H-3'), 3.69 (s, 6H,  $2\text{XOMe}$ ), 3.13 (dd,  $J_{5',5''}=11.0$ ,  $J_{5',4'}=4.6$ , 1H, H-5''), 0.82 (s, 9H, *t*-Bu), 0.08 (s, 3H, Me), 0.06 (s, 3H, Me), HRMS (FAB<sup>+</sup>) calcd for  $\text{C}_{42}\text{H}_{48}\text{N}_6\text{O}_6\text{S}_2\text{Si} (\text{MH}^+)$  825.2924, found 825.2977.

**2'-*t*-Butyldimethylsilyl-3'-deoxy-3'-S-pyridylsulfanyl-5'-O-(4,4'-dimethoxytrityl)-N<sup>2</sup>-isobutyrylguanosine (21).** To the solution of 20 (310 mg, 0.38 mmol) in dry pyridine (5 ml) isobutyric anhydride (0.19 ml, 1.14 mmol) and DMAP (46 mg, 0.38 mm) were added and the mixture was stirred at rt 16 h. It was then stirred at 50 °C for 5 h, quenched with methanol (2 ml) and evaporated to a syrup *in vacuo*. The residue was partitioned between dichloromethane and 5% aqueous  $\text{NaHCO}_3$ , the organic layer was washed with brine, dried ( $\text{Na}_2\text{SO}_4$ ) and evaporated to a syrup. Column chromatography on silica gel using 1-5% gradient of methanol in  $\text{CH}_2\text{Cl}_2$  afforded product as a colorless foam (320 mg, 95%), <sup>1</sup>H NMR δ 11.94 (br s, 1H, NH), 8.36 (m, 1H, Pyr), 7.85 (m, 1H, Pyr), 7.80 (s, 1H, H-8), 7.58-6.71 (m, 16H, NH, aromatic), 5.83 (d,  $J_{1',2'}=5.2$ , 1H, H-1'), 5.22 (app t,  $J_{1',2'}=5.2$ , 1H, H-2'), 4.50 (m, 1H, H-4'), 4.27 (app t,  $J_{3',4'}=6.4$ , 1H, H-3'), 3.76 (s, 3H, OMe), 3.75 (s, 3H, OMe), 3.56 (dd,  $J_{5',5''}=11.0$ ,  $J_{5',4'}=1.8$ , 1H, H5'), 2.98 (dd,  $J_{5',5''}=11.0$ ,  $J_{5',4'}=3.0$ , 1H, H-5''), 1.69 (m, 1H,  $\text{CHMe}_2$ ), 0.94 (d,  $J=7.2$ , 3H,  $\text{CH}_3$ ), 0.76 (d,  $J=7.2$ , 3H,  $\text{CH}_3$ ), 0.88 (s, 9H, *t*-Bu), 0.11 (s, 3H, Me), 0.06 (s, 3H, Me), HRMS (FAB<sup>+</sup>) calcd for  $\text{C}_{46}\text{H}_{54}\text{N}_6\text{O}_7\text{S}_2\text{Si} (\text{MH}^+)$  895.3343, found 895.3380.

**2'-O-t-Butyldimethylsilyl-3'-deoxy-3'-thio-5'-O-(4,4'-dimethoxytrityl)-N<sup>2</sup>-isobutyrylguanosine (22).** To the solution of 21 (340 mg, 0.38 mmol) in chloroform (20 ml) TEA (0.4 ml) and dithiotreitol DTT (140 mg, 0.91 mmol) were added and the reaction mixture was stirred for 1 h at rt. The reaction mixture was washed with saturated aqueous 5 NaHCO<sub>3</sub>, water, dried (Na<sub>2</sub>SO<sub>4</sub>) and concentrated to a syrup. Silica gel column chromatography using 0.5-2% gradient of methanol in CH<sub>2</sub>Cl<sub>2</sub> afforded 22 (270 mg, 90%), <sup>1</sup>H NMR for the major rotamer: δ 11.92 (br s, 1H, NH), 7.93 (s, 1H, H-8), 7.63-6.80 (m, 16H, NH, aromatic), 5.83 (d, J<sub>1',2'</sub>=2.8, 1H, H-1'), 4.74 (dd, J<sub>2',1'</sub>=2.8, J<sub>2',3'</sub>=5.4, 1H, H-2'), 4.13 (dd, J<sub>4',3'</sub>=7.6, J<sub>4',5'</sub>=1.2, 1H, H-4'), 3.78 (s, 3H, OMe), 3.73 (s, 3H, OMe), 10 3.63 (dd, J<sub>5',5'</sub>=11.0, J<sub>5',4'</sub>=1.2, 1H, HS'), 3.27 (dd, J<sub>5'',5'</sub>=11.0, J<sub>5'',4'</sub>=3.0, 1H, H-5''), 1.63 (d, J<sub>SH,3'</sub>=8.4, 1H, SH), 2.08 (m, 1H, CHMe<sub>2</sub>), 1.09 (d, J=6.8, 3H, CH<sub>3</sub>), 0.98 (d, J=6.8, 3H, CH<sub>3</sub>), 0.91 (s, 9H, t-Bu), 0.14 (s, 3H, Me), 0.08 (s, 3H, Me), HRMS (FAB<sup>+</sup>) calcd for C<sub>41</sub>H<sub>51</sub>N<sub>5</sub>O<sub>7</sub>SSi (MH<sup>+</sup>) 786.3357, found 786.3354.

**2'-O-t-Butyldimethylsilyl-3'-deoxy-3'-thio-5'-O-(4,4'-dimethoxytrityl)-N<sup>2</sup>-isobutyrylguanosine 3'-S-(2-cyanoethyl) N,N-diisopropylphoramidite (23).** Phosphitylation of 22 as described by Sun *et al.* afforded product which was purified by flash chromatography using 0.5% ethanol in CH<sub>2</sub>Cl<sub>2</sub> containing 1% TEA. Final product was obtained as a white powder by precipitation from toluene-pentane at 0 °C (76% yield), <sup>31</sup>P NMR δ 163.5 (s), 159.6 (s), HRMS (FAB<sup>+</sup>) calcd for C<sub>50</sub>H<sub>68</sub>N<sub>7</sub>O<sub>8</sub>PSSi (MH<sup>+</sup>) 986.4435, found 986.4406.

Example 13: Synthesis of 5'- thiophosphate nucleoside phosphoramidite and preparation of solid support

Referring to Figure 21, Applicant shows a scheme for synthesis of 5'-deoxy-5'-thionucleoside phosphoramidites and succinates. For example Lehmann *et al.* (NAR 25 1989, 17, 2379; incorporated by reference herein) describes the preparation of the more base-stable sarcosyl modified solid support.

Matulic-Adamic. *et al.* in *Nucleosides & Nucleosides* 1997, 16, 1933, describes the incorporation of 5'-thio modification into oligonucleotides. Applicant additionally describes the a method where, the cleavage of the 5'-protecting group on the solid support 30 in order to elongate the chain is accomplished by using, instead of AgNO<sub>3</sub> in CH<sub>3</sub>CN, 0.1M 1,8-diazabicyclo[5.4.0]undec-7-ene (DBU) in CH<sub>3</sub>CN or 20% piperidine in CH<sub>3</sub>CN.

Note that 5'S-DMT protection is cleaved with iodine (Henningfeld *et al.* *JACS* 1996, 118, 11701) thus complicating the oligo synthesis, while S-Fm is reported to be 35 resistant to 0.1 M iodine in DMF.

Oligonucleotides were synthesized with 5'-thiophosphate linkage and tested *in vitro* in a standard RNase H cleavage assay described in McSwiggen, US Patent No. 5,525,468; incorporated by reference herein.

**Example 14: Synthesis of 5'-O-dimethoxytrityl-3'-deoxy-3'-thio-3'-S-(2-Cyanoethyl-N,N-**

**Diisopropylphosphoramidite-2'-O-methyl uridine (6) (Figure 22)**

***5'-O-tert-butyldiphenylsilyl-2'-O-methyl uridine (1):*** To a solution of 2'-O-methyl uridine stirring at 0 C under positive pressure argon in anhydrous pyridine was added *tert*-butyldiphenylsilyl chloride (1.2 eq.) The reaction mixture was allowed to warm to rt and was maintained at rt for 18 hours at which time ethanol was added, pyridine removed in *vacuo*, and the reaction residue partitioned between dichloromethane and sat. aq. sodium bicarbonate. The organic layer was then dried over sodium sulfate. Flash chromatography using an ethyl acetate/hexanes gradient gave (1) as a white foam.

***5'-O-tert-butyldiphenylsilyl-2'-O-methyl-2,3'-anhydro uridine (2):*** To a solution of (1) and DEAD (3.5 eq.) stirring at 0 C under argon in anhydrous THF was added triphenylphosphine (3.5 eq.). The reaction mixture was warmed to rt and stirred at rt under argon for 18 hours, after which THF was removed *in vacuo*. The crude reaction residue was partitioned between dichloromethane and sat. aq. sodium bicarbonate, the organics dried over sodium sulfate preceding flash chromatography. An ethyl acetate/hexanes gradient afforded (2) as an off white foam.

***5'-O-tert-butyldiphenylsilyl-3'-S-acetyl-2'-O-methyl uridine (3):*** Compound (2) was treated with thiolacetic acid in dioxane at 100 C for 18 hours while stirring in a stainless steel bomb. The reaction mixture was evaporated *in vacuo* then purified by flash silica gel chromatography to give (3) as a light yellow foam.

***5'-O-dimethoxytrityl-3'-S-acetyl-2'-O-methyl uridine (4):*** To a solution of (3) stirring at rt under positive pressure argon was added 1M TBAF in THF buffered with acetic acid. The resulting clear, light yellow solution was stirred at rt for one hour, then THF removed *in vacuo*. Crude (4) was flash silica purified using an ethanol/dichloromethane gradient. The purified product was then co-evaporated with anhydrous pyridine, then dissolved in anhydrous pyridine. Dimethoxytrityl chloride was added to the reaction at rt and the resulting clear, reddish solution stirred at rt for 18 hours. Pyridine was removed in *vacuo* after quenching with ethanol, and the resulting crude foam partitioned between dichloromethane and sat. aq. sodium bicarbonate and the organics dried over sodium sulfate. Flash chromatography using an ethyl acetate/hexanes gradient furnished pure (4).

*5'-O-dimethoxytrityl-3'-deoxy-3'-thio-2'-O-methyl uridine (5):* Compound (4) was dissolved in 40% aq. methylamine in the presence of DTT. The reaction mixture was stirred at rt for one hour then evaporated *in vacuo*. Flash chromatography using an ethyl acetate/hexanes gradient gave (5) as an off white foam.

5       *5'-O-dimethoxytrityl-3'-deoxy-3'-thio-3'-S-(2-Cyanoethyl-N,N-Diisopropylphosphoramidite-2'-O-methyl uridine (6):* To a cooled (0 °C) solution of (5) and *N,N*-diisopropylethylamine in dry CH<sub>2</sub>Cl<sub>2</sub> was added 2-cyanoethyl *N,N*-diisopropylchlorophosphoramidite dropwise via syringe. The mixture was stirred at room temperature until all starting material was consumed (5 hr.) The reaction mixture was quenched with 10 anhydrous ethanol and diluted with hexanes. Flash chromatography using an ethyl acetate/hexanes gradient provided pure (6).

#### Diagnostic uses

Nucleic acid molecules of this invention may be used as diagnostic tools to examine genetic drift and mutations within diseased cells or to detect the presence of 15 specific RNAs in a cell. The close relationship between for example ribozyme activity and the structure of the target RNA allows the detection of mutations in any region of the molecule which alters the base-pairing and three-dimensional structure of the target RNA. By using multiple ribozymes described in this invention, one may map nucleotide changes which are important to RNA structure and function *in vitro*, as well as in cells and tissues. 20 Cleavage of target RNAs with ribozymes may be used to inhibit gene expression and define the role (essentially) of specified gene products in the progression of disease. In this manner, other genetic targets may be defined as important mediators of the disease. These experiments will lead to better treatment of the disease progression by affording the 25 possibility of combinational therapies (e.g., multiple nucleic acid molecules targeted to different genes, nucleic acid molecules coupled with known small molecule inhibitors, or intermittent treatment with combinations of nucleic acid molecules and/or other chemical or biological molecules). Other *in vitro* uses of nucleic acid molecules of this invention are well known in the art, and include detection of the presence of RNAs related to various 30 conditions. Such RNA is detected by determining the presence of a cleavage product after treatment with for example, an enzymatic nucleic acid molecule using standard methodology.

In a specific example, ribozymes which can cleave only wild-type or mutant forms 35 of the target RNA are used for the assay. The first ribozyme is used to identify wild-type RNA present in the sample and the second ribozyme will be used to identify mutant RNA in the sample. As reaction controls, synthetic substrates of both wild-type and mutant

RNA will be cleaved by both ribozymes to demonstrate the relative ribozyme efficiencies in the reactions and the absence of cleavage of the "non-targeted" RNA species. The cleavage products from the synthetic substrates will also serve to generate size markers for the analysis of wild-type and mutant RNAs in the sample population. Thus each analysis 5 will require two ribozymes, two substrates and one unknown sample which will be combined into six reactions. The presence of cleavage products will be determined using an RNase protection assay so that full-length and cleavage fragments of each RNA can be analyzed in one lane of a polyacrylamide gel. It is not absolutely required to quantify the results to gain insight into the expression of mutant RNAs and putative risk of the desired 10 phenotypic changes in target cells. The expression of mRNA whose protein product is implicated in the development of a phenotype is adequate to establish risk. If probes of comparable specific activity are used for both transcripts, then a qualitative comparison of RNA levels will be adequate and will decrease the cost of the initial diagnosis. Higher mutant form to wild-type ratios will be correlated with higher risk whether RNA levels are 15 compared qualitatively or quantitatively.

#### Additional Uses

Potential usefulness of sequence-specific enzymatic nucleic acid molecules of the instant invention might have many of the same applications for the study of RNA that DNA restriction endonucleases have for the study of DNA (Nathans *et al.*, 1975 *Ann. Rev. 20 Biochem.* 44:273). For example, the pattern of restriction fragments could be used to establish sequence relationships between two related RNAs, and large RNAs could be specifically cleaved to fragments of a size more useful for study. The ability to engineer sequence specificity of the ribozyme is ideal for cleavage of RNAs of unknown sequence.

Other embodiments are within the following claims.

TABLE ICharacteristics of naturally occurring ribozymesGroup I Introns

- Size: ~150 to >1000 nucleotides.
- 5 • Requires a U in the target sequence immediately 5' of the cleavage site.
- Binds 4-6 nucleotides at the 5'-side of the cleavage site.
- Reaction mechanism: attack by the 3'-OH of guanosine to generate cleavage products with 3'-OH and 5'-guanosine.
- 10 • Additional protein cofactors required in some cases to help folding and maintainance of the active structure.
- Over 300 known members of this class. Found as an intervening sequence in *Tetrahymena thermophila* rRNA, fungal mitochondria, chloroplasts, phage T4, blue-green algae, and others.
- Major structural features largely established through phylogenetic comparisons, mutagenesis, and biochemical studies [<sup>1,2</sup>].
- 15 • Complete kinetic framework established for one ribozyme [<sup>3,4,5,6</sup>].
- Studies of ribozyme folding and substrate docking underway [<sup>7,8,9</sup>].

<sup>1</sup> Michel, Francois; Westhof, Eric. Slippery substrates. *Nat. Struct. Biol.* (1994), 1(1), 5-7.

<sup>2</sup> Lisacek, Frederique; Diaz, Yolande; Michel, Francois. Automatic identification of group I intron cores in genomic DNA sequences. *J. Mol. Biol.* (1994), 235(4), 1206-17.

<sup>3</sup> Herschlag, Daniel; Cech, Thomas R.. Catalysis of RNA cleavage by the *Tetrahymena thermophila* ribozyme. 1. Kinetic description of the reaction of an RNA substrate complementary to the active site. *Biochemistry* (1990), 29(44), 10159-71.

<sup>4</sup> Herschlag, Daniel; Cech, Thomas R.. Catalysis of RNA cleavage by the *Tetrahymena thermophila* ribozyme. 2. Kinetic description of the reaction of an RNA substrate that forms a mismatch at the active site. *Biochemistry* (1990), 29(44), 10172-80.

<sup>5</sup> Knitt, Deborah S.; Herschlag, Daniel. pH Dependencies of the *Tetrahymena* Ribozyme Reveal an Unconventional Origin of an Apparent pKa. *Biochemistry* (1996), 35(5), 1560-70.

<sup>6</sup> Bevilacqua, Philip C.; Sugimoto, Naoki; Turner, Douglas H.. A mechanistic framework for the second step of splicing catalyzed by the *Tetrahymena* ribozyme. *Biochemistry* (1996), 35(2), 648-58.

<sup>7</sup> Li, Yi; Bevilacqua, Philip C.; Mathews, David; Turner, Douglas H.. Thermodynamic and activation parameters for binding of a pyrene-labeled substrate by the *Tetrahymena* ribozyme: docking is not diffusion-controlled and is driven by a favorable entropy change. *Biochemistry* (1995), 34(44), 14394-9.

<sup>8</sup> Banerjee, Aloke Raj; Turner, Douglas H.. The time dependence of chemical modification reveals slow steps in the folding of a group I ribozyme. *Biochemistry* (1995), 34(19), 6504-12.

- Chemical modification investigation of important residues well established [<sup>10, 11</sup>].
- The small (4-6 nt) binding site may make this ribozyme too non-specific for targeted RNA cleavage, however, the Tetrahymena group I intron has been used to repair a "defective" β-galactosidase message by the ligation of new β-galactosidase sequences onto the defective message [<sup>12</sup>].

5

**RNase P RNA (M1 RNA)**

- Size: ~290 to 400 nucleotides.
- RNA portion of a ubiquitous ribonucleoprotein enzyme.
- Cleaves tRNA precursors to form mature tRNA [<sup>13</sup>].
- 10 • Reaction mechanism: possible attack by M<sup>2+</sup>-OH to generate cleavage products with 3'-OH and 5'-phosphate.
- RNase P is found throughout the prokaryotes and eukaryotes. The RNA subunit has been sequenced from bacteria, yeast, rodents, and primates.
- Recruitment of endogenous RNase P for therapeutic applications is possible through 15 hybridization of an External Guide Sequence (EGS) to the target RNA [<sup>14, 15</sup>]
- Important phosphate and 2' OH contacts recently identified [<sup>16, 17</sup>]

---

<sup>9</sup> Zarrinkar, Patrick P.; Williamson, James R.. The P9.1-P9.2 peripheral extension helps guide folding of the Tetrahymena ribozyme. *Nucleic Acids Res.* (1996), 24(5), 854-8.

<sup>10</sup> Strobel, Scott A.; Cech, Thomas R.. Minor groove recognition of the conserved G.cndot.U pair at the Tetrahymena ribozyme reaction site. *Science (Washington, D. C.)* (1995), 267(5198), 675-9.

<sup>11</sup> Strobel, Scott A.; Cech, Thomas R.. Exocyclic Amine of the Conserved G.cndot.U Pair at the Cleavage Site of the Tetrahymena Ribozyme Contributes to 5'-Splice Site Selection and Transition State Stabilization. *Biochemistry* (1996), 35(4), 1201-11.

<sup>12</sup> Sullenger, Bruce A.; Cech, Thomas R.. Ribozyme-mediated repair of defective mRNA by targeted trans-splicing. *Nature (London)* (1994), 371(6498), 619-22.

<sup>13</sup> Robertson, H.D.; Altman, S.; Smith, J.D. *J. Biol. Chem.*, 247, 5243-5251 (1972).

<sup>14</sup> Forster, Anthony C.; Altman, Sidney. External guide sequences for an RNA enzyme. *Science (Washington, D. C.)*, 1883- (1990), 249(4970), 783-6.

<sup>15</sup> Yuan, Y.; Hwang, E. S.; Altman, S. Targeted cleavage of mRNA by human RNase P. *Proc. Natl. Acad. Sci. USA* (1992) 89, 8006-10.

<sup>16</sup> Harris, Michael E.; Pace, Norman R.. Identification of phosphates involved in catalysis by the ribozyme RNase P RNA. *RNA* (1995), 1(2), 210-18.

<sup>17</sup> Pan, Tao; Loria, Andrew; Zhong, Kun. Probing of tertiary interactions in RNA: 2'-hydroxyl-base contacts between the RNase P RNA and pre-tRNA. *Proc. Natl. Acad. Sci. U. S. A.* (1995), 92(26), 12510-14.

Group II Introns

- Size: >1000 nucleotides.
- Trans cleavage of target RNAs recently demonstrated [<sup>18, 19</sup>].
- Sequence requirements not fully determined.
- 5 • Reaction mechanism: 2'-OH of an internal adenosine generates cleavage products with 3'-OH and a "lariat" RNA containing a 3'-5' and a 2'-5' branch point.
- Only natural ribozyme with demonstrated participation in DNA cleavage [<sup>20, 21</sup>] in addition to RNA cleavage and ligation.
- Major structural features largely established through phylogenetic comparisons [<sup>22</sup>].
- 10 • Important 2' OH contacts beginning to be identified [<sup>23</sup>]
- Kinetic framework under development [<sup>24</sup>]

Neurospora VS RNA

- Size: ~144 nucleotides.
- TRANS CLEAVAGE OF HAIRPIN TARGET RNAs RECENTLY DEMONSTRATED [<sup>25</sup>].
- 15 • Sequence requirements not fully determined.
- Reaction mechanism: attack by 2'-OH 5' to the scissile bond to generate cleavage products with 2',3'-cyclic phosphate and 5'-OH ends.

<sup>18</sup> Pyle, Anna Marie; Green, Justin B.. Building a Kinetic Framework for Group II Intron Ribozyme Activity: Quantitation of Interdomain Binding and Reaction Rate. *Biochemistry* (1994), 33(9), 2716-25.

<sup>19</sup> Michels, William J. Jr.; Pyle, Anna Marie. Conversion of a Group II Intron into a New Multiple-Turnover Ribozyme that Selectively Cleaves Oligonucleotides: Elucidation of Reaction Mechanism and Structure/Function Relationships. *Biochemistry* (1995), 34(9), 2965-77.

<sup>20</sup> Zimmerly, Steven; Guo, Huatao; Eskes, Robert; Yang, Jian; Perlman, Philip S.; Lambowitz, Alan M.. A group II intron RNA is a catalytic component of a DNA endonuclease involved in intron mobility. *Cell* (Cambridge, Mass.) (1995), 83(4), 529-38.

<sup>21</sup> Griffin, Edmund A., Jr.; Qin, Zhifeng; Michels, Williams J., Jr.; Pyle, Anna Marie. Group II intron ribozymes that cleave DNA and RNA linkages with similar efficiency, and lack contacts with substrate 2'-hydroxyl groups. *Chem. Biol.* (1995), 2(11), 761-70.

<sup>22</sup> Michel, Francois; Ferat, Jean Luc. Structure and activities of group II introns. *Annu. Rev. Biochem.* (1995), 64, 435-61.

<sup>23</sup> Abramovitz, Dana L.; Friedman, Richard A.; Pyle, Anna Marie. Catalytic role of 2'-hydroxyl groups within a group II intron active site. *Science* (Washington, D. C.) (1996), 271(5254), 1410-13.

<sup>24</sup> Daniels, Danette L.; Michels, William J., Jr.; Pyle, Anna Marie. Two competing pathways for self-splicing by group II introns: a quantitative analysis of in vitro reaction rates and products. *J. Mol. Biol.* (1996), 256(1), 31-49.

<sup>25</sup> Guo, Hans C. T.; Collins, Richard A.. Efficient trans-cleavage of a stem-loop RNA substrate by a ribozyme derived from Neurospora VS RNA. *EMBO J.* (1995), 14(2), 368-76.

- Binding sites and structural requirements not fully determined.
- Only 1 known member of this class. Found in *Neurospora* VS RNA.

### Hammerhead Ribozyme

(see text for references)

5     • Size: ~13 to 40 nucleotides.

• Requires the target sequence UH immediately 5' of the cleavage site.

• Binds a variable number nucleotides on both sides of the cleavage site.

• Reaction mechanism: attack by 2'-OH 5' to the scissile bond to generate cleavage products with 2',3'-cyclic phosphate and 5'-OH ends.

10    • 14 known members of this class. Found in a number of plant pathogens (virusoids) that use RNA as the infectious agent.

• Essential structural features largely defined, including 2 crystal structures [<sup>26,27</sup>]

• Minimal ligation activity demonstrated (for engineering through *in vitro* selection) [<sup>28</sup>]

• Complete kinetic framework established for two or more ribozymes [<sup>29</sup>].

15    • Chemical modification investigation of important residues well established [<sup>30</sup>].

### Hairpin Ribozyme

• Size: ~50 nucleotides.

• Requires the target sequence GUC immediately 3' of the cleavage site.

• Binds 4-6 nucleotides at the 5'-side of the cleavage site and a variable number to the 3'-side of the cleavage site.

20    • Reaction mechanism: attack by 2'-OH 5' to the scissile bond to generate cleavage products with 2',3'-cyclic phosphate and 5'-OH ends.

---

<sup>26</sup> Scott, W.G., Finch, J.T., Aaron, K. The crystal structure of an all RNA hammerhead ribozyme: A proposed mechanism for RNA catalytic cleavage. *Cell*, (1995), 81, 991-1002.

<sup>27</sup> McKay, Structure and function of the hammerhead ribozyme: an unfinished story. *RNA*, (1996), 2, 395-403.

<sup>28</sup> Long, D., Uhlenbeck, O., Hertel, K. Ligation with hammerhead ribozymes. US Patent No. 5,633,133.

<sup>29</sup> Hertel, K.J., Herschlag, D., Uhlenbeck, O. A kinetic and thermodynamic framework for the hammerhead ribozyme reaction. *Biochemistry*, (1994) 33, 3374-3385. Beigelman, L., *et al.*, Chemical modifications of hammerhead ribozymes. *J. Biol. Chem.*, (1995) 270, 25702-25708.

<sup>30</sup> Beigelman, L., *et al.*, Chemical modifications of hammerhead ribozymes. *J. Biol. Chem.*, (1995) 270, 25702-25708.

- 3 known members of this class. Found in three plant pathogen (satellite RNAs of the tobacco ringspot virus, arabis mosaic virus and chicory yellow mottle virus) which uses RNA as the infectious agent.
- Essential structural features largely defined [<sup>31, 32, 33, 34</sup>]
- 5 • Ligation activity (in addition to cleavage activity) makes ribozyme amenable to engineering through *in vitro* selection [<sup>35</sup>]
- Complete kinetic framework established for one ribozyme [<sup>36</sup>].
- Chemical modification investigation of important residues begun [<sup>37, 38</sup>].

#### Hepatitis Delta Virus (HDV) Ribozyme

- 10 • Size: ~60 nucleotides.
- Trans cleavage of target RNAs demonstrated [<sup>39</sup>].
- Binding sites and structural requirements not fully determined, although no sequences 5' of cleavage site are required. Folded ribozyme contains a pseudoknot structure [<sup>40</sup>].

<sup>31</sup> Hampel, Arnold; Tritz, Richard; Hicks, Margaret; Cruz, Phillip. 'Hairpin' catalytic RNA model: evidence for helices and sequence requirement for substrate RNA. *Nucleic Acids Res.* (1990), 18(2), 299-304.

<sup>32</sup> Chowrira, Bharat M.; Berzal-Herranz, Alfredo; Burke, John M.. Novel guanosine requirement for catalysis by the hairpin ribozyme. *Nature (London)* (1991), 354(6351), 320-2.

<sup>33</sup> Berzal-Herranz, Alfredo; Joseph, Simpson; Chowrira, Bharat M.; Butcher, Samuel E.; Burke, John M.. Essential nucleotide sequences and secondary structure elements of the hairpin ribozyme. *EMBO J.* (1993), 12(6), 2567-73.

<sup>34</sup> Joseph, Simpson; Berzal-Herranz, Alfredo; Chowrira, Bharat M.; Butcher, Samuel E.. Substrate selection rules for the hairpin ribozyme determined by *in vitro* selection, mutation, and analysis of mismatched substrates. *Genes Dev.* (1993), 7(1), 130-8.

<sup>35</sup> Berzal-Herranz, Alfredo; Joseph, Simpson; Burke, John M.. *In vitro* selection of active hairpin ribozymes by sequential RNA-catalyzed cleavage and ligation reactions. *Genes Dev.* (1992), 6(1), 129-34.

<sup>36</sup> Hegg, Lisa A.; Fedor, Martha J.. Kinetics and Thermodynamics of Intermolecular Catalysis by Hairpin Ribozymes. *Biochemistry* (1995), 34(48), 15813-28.

<sup>37</sup> Grasby, Jane A.; Mersmann, Karin; Singh, Mohinder; Gait, Michael J.. Purine Functional Groups in Essential Residues of the Hairpin Ribozyme Required for Catalytic Cleavage of RNA. *Biochemistry* (1995), 34(12), 4068-76.

<sup>38</sup> Schmidt, Sabine; Beigelman, Leonid; Karpeisky, Alexander; Usman, Nassim; Sorensen, Ulrik S.; Gait, Michael J.. Base and sugar requirements for RNA cleavage of essential nucleoside residues in internal loop B of the hairpin ribozyme: implications for secondary structure. *Nucleic Acids Res.* (1996), 24(4), 573-81.

<sup>39</sup> Perrotta, Anne T.; Been, Michael D.. Cleavage of oligoribonucleotides by a ribozyme derived from the hepatitis delta virus RNA sequence. *Biochemistry* (1992), 31(1), 16-21.

<sup>40</sup> Perrotta, Anne T.; Been, Michael D.. A pseudoknot-like structure required for efficient self-cleavage of hepatitis delta virus RNA. *Nature (London)* (1991), 350(6317), 434-6.

- Reaction mechanism: attack by 2'-OH 5' to the scissile bond to generate cleavage products with 2',3'-cyclic phosphate and 5'-OH ends.
- Only 2 known members of this class. Found in human HDV.
- Circular form of HDV is active and shows increased nuclease stability [<sup>41</sup>]

---

<sup>41</sup> Puttaraju, M.; Perrotta, Anne T.; Been, Michael D.. A circular trans-acting hepatitis delta virus ribozyme. *Nucleic Acids Res.* (1993), 21(18), 4253-8.

**Table II: 2.5  $\mu$ mol RNA Synthesis Cycle**

Reagent	Equivalents	Amount	Wait Time*
Phosphoramidites	6.5	163 $\mu$ L	2.5
<i>S</i> -Ethyl Tetrazole	23.8	238 $\mu$ L	2.5
Acetic Anhydride	100	233 $\mu$ L	5 sec
<i>N</i> -Methyl Imidazole	186	233 $\mu$ L	5 sec
TCA	83.2	1.73 mL	21 sec
Iodine	8.0	1.18 mL	45 sec
Acetonitrile	NA	6.67 mL	NA

Table IIIA. Antisense Sequences and Corresponding in vitro Data

Seq. I.D. No.	Target	Cell Line	Sequence	Inhibition
2717	c-raf	PC-3	u,c,c,ccgc[C,T,G,T,G,A,C]augc,a,u,T	70-80%
2718	Estrogen Receptor (ER)	MCF-7	a,ggc,auccaa[C,A,A,G,G,C,A,i]cugacc,a,u,c	53%
2719	Estrogen Receptor (ER)	MCF-7	c,a,g,caucca[A,C,A,A,G,G,C,C]acugac,c,a,u	48%
2720	Estrogen Receptor (ER)	MCF-7	c,u,g,ccagggt[T,G,G,T,C,A,G,]uaaggcc,c,s,u	60%
2721	Estrogen Receptor (ER)	MCF-7	u,u,u,cccugggt[T,C,C,T,G,T,]ccaaga,g,s,c,a	59%
2722	Estrogen Receptor (ER)	MCF-7	c,s,c,a,gcauc[T,C,C,A,G,C,A,s]gcagg,u,c,a	61%
2723	Estrogen Receptor (ER)	MCF-7	c,g,u,ccaggc[A,T,C,T,C,C,A,i]gcagg,c,a,g,g	64%

Table III B. Ribozyme sequences and Corresponding *in vitro* Data

Seq. I.D. No.	Target	Cell Line	Sequence	Inhibition
2724	Estrogen Receptor (ER)	MCF-7	[A <sub>s</sub> T <sub>s</sub> A <sub>s</sub> G <sub>s</sub> A <sub>s</sub> T <sub>s</sub> T <sub>s</sub> T <sub>s</sub> ] cUGAU <del>G</del> aggccgaaaggccGaa Aggcacac B	50%
2725	Estrogen Receptor (ER)	MCF-7	g <sub>s</sub> g <sub>s</sub> u <sub>s</sub> c <sub>s</sub> agu cUGAU <del>G</del> aggccg <del>u</del> uaggccGaa Agc <del>C</del> CC [A <sub>s</sub> T <sub>s</sub> C <sub>s</sub> A <sub>s</sub> T <sub>s</sub> C <sub>s</sub> G] B	see figure 4
2726	Control	MCF-7	g <sub>s</sub> u <sub>s</sub> C <sub>s</sub> ggcc cUuAu <del>G</del> aggccg <del>u</del> uaggccGau Acagu[T <sub>s</sub> A <sub>s</sub> G <sub>s</sub> C <sub>s</sub> T <sub>s</sub> A] B	See figure 4

Lower case = 2'OMe

U = 2'-C-Allyl-U

G,A= ribo G,A

s = phosphorothioate linkages

B = inverted abasic

[G,A,C,T]= DNA

Table IV. Hammerhead Ribozyme and Target Sequences For Estrogen Receptor

Pos	RZ	Seq. ID. No.	Substrate	Seq. ID. No.
5	22 UUGGCUUA CUGAUGAG X CGAA ACAUCACU	1	AGTGATGTT TAAGCCAA	1246
	23 AUUGGCUU CUGAUGAG X CGAA AACAUAC	2	GTGATGTT AAGCCAAT	1247
	24 CAUUGGCU CUGAUGAG X CGAA AAACAUCA	3	TGATGTTT A AGCCAATG	1248
	34 CUUGCCU CUGAUGAG X CGAA ACAUUGGC	4	GCCAATGT C AGGGCAAG	1249
10	51 CGGCCAGG CUGAUGAG X CGAA ACUGUUGC	5	GCAACAGT C CCTGGCCG	1250
	61 UGCUGGGAG CUGAUGAG X CGAA ACGGCCAG	6	CTGGCCGT C CTCCAGCA	1251
	64 AGGUGCUG CUGAUGAG X CGAA AGGACGGC	7	GCCGTCCT C CAGCACCT	1252
	73 GCAUUACA CUGAUGAG X CGAA AGGUGCUG	8	CAGCACCT T TGTAATGC	1253
15	74 UGCAUUAC CUGAUGAG X CGAA AAGGUGCU	9	AGCACCTT T GTAATGCA	1254
	77 AUAUGCAU CUGAUGAG X CGAA ACAAAGGU	10	ACCTTTGT A ATGCATAT	1255
	84 CGAGCUCA CUGAUGAG X CGAA AUGCAUUA	11	TAATGCAT A TGAGCTCG	1256
	91 GGUCUCCC CUGAUGAG X CGAA AGCUCAUA	12	TATGAGCT C GGGAGACC	1257
20	103 ACUUUAAG CUGAUGAG X CGAA ACUGGUCU	13	AGACCAAGT A CTTAAAGT	1258
	106 CCAACUUU CUGAUGAG X CGAA AGUACUGG	14	CCAGTACT T AAAGTTGG	1259
	107 UCCAACUU CUGAUGAG X CGAA AAGUACUG	15	CAGTACTT A AAGTTGGA	1260
	112 GGGCCUCC CUGAUGAG X CGAA ACUUUAAG	16	CTTAAAGT T GGAGGCC	1261
25	148 CCAGGACG CUGAUGAG X CGAA ACGCCCUC	17	GAGGGCGT T CGTCCTGG	1262
	149 CCCAGGAC CUGAUGAG X CGAA AACGCCU	18	AGGGCGTT C GTCCTGGG	1263
	152 GCUCCCCAG CUGAUGAG X CGAA ACGAACGC	19	CGCTTCGT C CTGGGAGC	1264
	167 GACGGAGC CUGAUGAG X CGAA AGUGCAGC	20	GCTGCACT T GCTCCGTC	1265
	171 ACCCGACG CUGAUGAG X CGAA AGCAAGUG	21	CACTTGCT C CGTCGGGT	1266
30	175 GGGGACCC CUGAUGAG X CGAA ACGGAGCA	22	TGCTCCGT C GGGTCGCC	1267
	180 AAGCCGGC CUGAUGAG X CGAA ACCCGACG	23	CGTCGGGT C GCCGGCTT	1268
	188 GUCCGGUG CUGAUGAG X CGAA AGCCGGCG	24	CGCCGGCT T CACCGGAC	1269
	189 GGUCCGGU CUGAUGAG X CGAA AAGCCGGC	25	GCCGGCTT C ACCGGACC	1270
	205 UGCCCCGG CUGAUGAG X CGAA AGCCUGCG	26	CGCAGGCT C CCGGGGCA	1271
	231 CGACACGC CUGAUGAG X CGAA AGCUCUGG	27	CCAGAGCT C GCGTGTG	1272

238	GUCCCGCC CUGAUGAG X CGAA ACACCGA	28	TCGCGTGT C GGCGGGAC	1273
258	UUAGAGGC CUGAUGAG X CGAA ACGCAGCG	29	CGCTGCGT C GCCTCTAA	1274
263	CGAGGUUA CUGAUGAG X CGAA AGGGGACG	30	CGTCGCCT C TAACCTCG	1275
265	CCCGAGGU CUGAUGAG X CGAA AGAGGCGA	31	TCGCCTCT A ACCTCGGG	1276
270	CACAGCCC CUGAUGAG X CGAA AGGUUAGA	32	TCTAACCT C GGGCTGTG	1277
281	UGGAAAAA CUGAUGAG X CGAA AGCACAGC	33	OCTGTGCT C TTTTCCA	1278
283	CCUGGAAA CUGAUGAG X CGAA AGAGCACA	34	TGTGCTCT T TTTCCAGG	1279
284	ACCUGGAA CUGAUGAG X CGAA AAGAGCAC	35	GTGCTCTT T TTCCAGGT	1280
285	CACCUUGGA CUGAUGAG X CGAA AAAGAGCA	36	TGCTCTTT T TCCAGGTG	1281
286	CCACCUGG CUGAUGAG X CGAA AAAAGAGC	37	GCTCTTTT T CCAGGTGG	1282
287	GCCACCUUG CUGAUGAG X CGAA AAAAAGAG	38	CTCTTTT C CAGGTGGC	1283
304	GGCUCAGA CUGAUGAG X CGAA ACCGGCGG	39	CCGCGGGT T TCTGAGCC	1284
305	AGGCUCAG CUGAUGAG X CGAA ACCGGCG	40	CGCCGGTT T CTGAGCCT	1285
306	AAGGCUCA CUGAUGAG X CGAA AAACCGGC	41	GCCGGTTT C TGAGCCTT	1286
314	CAGGGCAG CUGAUGAG X CGAA AGGCUCAG	42	CTGAGCCT T CTGCCCTG	1287
315	GCAGGGCA CUGAUGAG X CGAA AAGGCUCA	43	TGAGCCTT C TGCCCTGC	1288
335	AGGGUGCA CUGAUGAG X CGAA ACCGUGUC	44	GACACGGT C TGCACCCCT	1289
375	UUGGUGUG CUGAUGAG X CGAA AGGGUCAU	45	ATGACCT C CACACCAA	1290
389	CCAUCCCCA CUGAUGAG X CGAA AUGCUUUG	46	CAAAGCAT C TGGGATGG	1291
402	UGAUGCAG CUGAUGAG X CGAA AGGGCCAU	47	ATGGCCCT A CTGCATCA	1292
409	UUGGAUCU CUGAUGAG X CGAA AUGCAGUA	48	TACTGCAT C AGATCCAA	1293
414	UUCCCCUUG CUGAUGAG X CGAA AUCUGAUG	49	CATCAGAT C CAAGGGAA	1294
445	GAGCUGCG CUGAUGAG X CGAA ACGGUUCA	50	TGAACCGT C CGCAGCTC	1295
453	GGGAUCUU CUGAUGAG X CGAA AGCUGCGG	51	CCGAGCT C AAGATCCC	1296
459	UCCAGGGG CUGAUGAG X CGAA AUCUUGAG	52	CTCAAGAT C CCCCTGGA	1297
488	UGUCCAGG CUGAUGAG X CGAA ACACCUUG	53	CGAGGTGT A CCTGGACA	1298
515	GGUAGUUG CUGAUGAG X CGAA ACACGGCG	54	CGCCGTGT A CAACTACC	1299
521	CCUCGGGG CUGAUGAG X CGAA AGUUGUAC	55	GTACAACAT A CCCCCGAGG	1300
539	UGAACUCG CUGAUGAG X CGAA AGGGGGCG	56	CGCCGCCT A CGAGTTCA	1301
545	CGGGCGUUG CUGAUGAG X CGAA ACUCGUAG	57	CTACGAGT T CAACGCCG	1302
546	GGGGCGUU CUGAUGAG X CGAA AACUCGUA	58	TACGAGTT C AACGCCGC	1303

576	UGACCGUA CUGAUGAG X CGAA ACCUGCGC	59	GCGCAGGT C TACGGTCA	1304
578	UCUGACCG CUGAUGAG X CGAA AGACCUGC	60	GCAGGTCT A CGGTAGA	1305
583	GCCGGUCU CUGAUGAG X CGAA ACCGUAGA	61	TCTACGGT C AGACCGGC	1306
594	CCGUAGGG CUGAUGAG X CGAA AGGCCGGU	62	ACCGGCCT C CCCTACGG	1307
599	CGGGGCCG CUGAUGAG X CGAA AGGGGAGG	63	CCTCCCT A CGGCCCCG	1308
611	CAGCCUCA CUGAUGAG X CGAA ACCCGGGG	64	CCCCGGGT C TGAGGCTG	1309
626	UGGAGCCG CUGAUGAG X CGAA ACGCCGCA	65	TGCGGCGT T CGGCTCCA	1310
627	UUGGAGCC CUGAUGAG X CGAA AACGCCGC	66	GCGGCGTT C GGCTCCAA	1311
632	GGCCGUUG CUGAUGAG X CGAA AGCCGAAC	67	GTTCGGCT C CAACGGCC	1312
649	UGGGGGGA CUGAUGAG X CGAA ACCCCCCA	68	TGGGGGGTT T TCCCCCCC	1313
650	GUGGGGGG CUGAUGAG X CGAA AACCCCCC	69	GGGGGGTT T CCCCCCAC	1314
651	AGUGGGGG CUGAUGAG X CGAA AAACCCCC	70	GGGGGTTT C CCCCCACT	1315
660	ACGCUGUU CUGAUGAG X CGAA AGUGGGGG	71	CCCCCACT C AACAGCGT	1316
671	GGCUCCGA CUGAUGAG X CGAA ACACCGUG	72	CAGCGTGT C TCCGAGCC	1317
673	CGGGCUCG CUGAUGAG X CGAA AGACACGC	73	GCGTGTCT C CGAGCCCG	1318
690	GGGUGCAG CUGAUGAG X CGAA AGCAUCAG	74	CTGATGCT A CTGCACCC	1319
713	GGAAAGGC CUGAUGAG X CGAA ACAGCUGC	75	GCAGCTGT C GCCTTTC	1320
718	CUGCAGGA CUGAUGAG X CGAA AGGCGACA	76	TGTCGCCT T TCCTGCAG	1321
719	GCUGCAGG CUGAUGAG X CGAA AAGGGCAG	77	GTCGCCTT T CCTGCAGC	1322
720	GGCUGCGAG CUGAUGAG X CGAA AAAGGCAG	78	TCGCCTTT C CTGCAGCC	1323
749	CCAGGUAG CUGAUGAG X CGAA AGGGCACC	79	GGTGCCT A CTACCTGG	1324
752	UCUCCAGG CUGAUGAG X CGAA AGUAGGGC	80	GCCCTACT A CCTGGAGA	1325
776	GCACCGUG CUGAUGAG X CGAA AGCCGUG	81	CAGCGGCT A CACGGTGC	1326
806	GCCUGUAG CUGAUGAG X CGAA AUGCCGGC	82	GCCGGCAT T CTACAGGC	1327
807	GGCCUGUA CUGAUGAG X CGAA AAUGCCGG	83	CCGGCATT C TACAGGCC	1328
809	UUGGCCUG CUGAUGAG X CGAA AGAAUGCC	84	GGCATTCT A CAGGCCAA	1329
820	AUUAUCUG CUGAUGAG X CGAA AUUUGGCC	85	GGCCAAATT CAGATAAT	1330
821	GAUUAUCU CUGAUGAG X CGAA AAUUUGGC	86	GCCAAATT C AGATAATC	1331
826	GCGUCGAU CUGAUGAG X CGAA AUCUGAAU	87	ATTCAGAT A ATCGACGC	1332
829	CUGGCGUC CUGAUGAG X CGAA AUUAUCUG	88	CAGATAAT C GACGCCAG	1333
854	UACUGGCC CUGAUGAG X CGAA AUCUUUCU	89	AGAAAGATT GGCCAGTA	1334

862	GUCAUUGG CUGAUGAG X CGAA ACUGGCCA	90	TGCCAGT A CCAATGAC	1335
880	CAUAGCCA CUGAUGAG X CGAA ACUUCCU	91	AGGGAAGT A TGGCTATG	1336
886	AGAUUCCA CUGAUGAG X CGAA AGCCAUAC	92	GTATGGCT A TGGAATCT	1337
893	CCUUGGCA CUGAUGAG X CGAA AUUCCAU	93	TATGGAAT C TGCCAAGG	1338
907	ACAGUAGC CUGAUGAG X CGAA AGUCUCCU	94	AGGAGACT C GCTACTGT	1339
911	CUGCACAG CUGAUGAG X CGAA AGCGAGUC	95	GAATCGCT A CTGTGCAG	1340
932	CUGAAGCA CUGAUGAG X CGAA AGUCAUUG	96	CAATGACT A TGCTTCAG	1341
937	GUAGCCUG CUGAUGAG X CGAA AGCAUAGU	97	ACTATGCT T CAGGCTAC	1342
938	GGUAGCCU CUGAUGAG X CGAA AAGCAUAG	98	CTATGCTT C AGGCTACC	1343
944	CAUAAUGG CUGAUGAG X CGAA AGCCUGAA	99	TTCAGGCT A CCATTATG	1344
949	GACUCCAU CUGAUGAG X CGAA AUGGUAGC	100	GCTACCAT T ATGGAGTC	1345
950	AGACUCCA CUGAUGAG X CGAA AAUGGUAG	101	CTACCAATT A TGGAGTCT	1346
957	CAGGACCA CUGAUGAG X CGAA ACUCCAU	102	TATGGAGT C TGGTCCTG	1347
962	CCUCACAG CUGAUGAG X CGAA ACCAGACU	103	AGTCTGGT C CTGTGAGG	1348
983	UCUUGAAG CUGAUGAG X CGAA AGGCCUUG	104	CAAGGCCT T TTCAAGA	1349
984	CUCUUGAA CUGAUGAG X CGAA AAGGCCUU	105	AAGGCCTT C TTCAAGAG	1350
986	UUCUCUUG CUGAUGAG X CGAA AGAAGGCC	106	GGCCTTCT T CAAGAGAA	1351
987	CUUCUCUU CUGAUGAG X CGAA AAGAAGGC	107	GCCTTCTT C AAGAGAAG	1352
997	UCCUUGAA CUGAUGAG X CGAA ACUUUCU	108	AGAGAAGT A TTCAAGGA	1353
999	UGUCCUUG CUGAUGAG X CGAA AUACUUCU	109	AGAAGTATT C AAGGACA	1354
1000	AUGUCCUU CUGAUGAG X CGAA AAUACUUC	110	GAAGTATT C AAGGACAT	1355
1009	AUAGUCGU CUGAUGAG X CGAA AUGUCCUU	111	AAGGACAT A ACGACTAT	1356
1016	GACACAU A CUGAUGAG X CGAA AGUCGUUA	112	TAACGACT A TATGTGTC	1357
1018	UGGACACA CUGAUGAG X CGAA AUAGUCGU	113	ACGACTAT A TGTGTCCA	1358
1024	GGUGGCUG CUGAUGAG X CGAA ACACAU	114	ATATGTGT C CAGCCACC	1359
1047	UUUUUAUC CUGAUGAG X CGAA AUGGUGCA	115	TGCACCAT T GATAAAAA	1360
1051	CCUGUUUU CUGAUGAG X CGAA AUCAAUGG	116	CCATTGAT A AAAACAGG	1361
1086	CAUUUGCG CUGAUGAG X CGAA AGCCGGCA	117	TGCCGGCT C CGCAAATG	1362
1097	CCACUUCG CUGAUGAG X CGAA AGCAUUUG	118	CAAATGCT A CGAAGTGG	1363
1125	UCUUUUUCG CUGAUGAG X CGAA AUCCCACC	119	GGTGGGAT A CGAAAAGA	1364
1154	UGUGUUUC CUGAUGAG X CGAA ACAUUCUC	120	GAGAATGTT GAAACACA	1365

1205	CUCCAGCA CUGAUGAG X CGAA ACCCCACU	121	AGTGGGGT C TGCTGGAG	1366
1233	CUUGGCCA CUGAUGAG X CGAA AGGUUGGC	122	GCCAACCT T TGGCCAAG	1367
1234	GCUUUGGCC CUGAUGAG X CGAA AAGGUUGG	123	CCAACCTT T GGCCAAGC	1368
1248	UUGAUCAU CUGAUGAG X CGAA AGCGGGCU	124	AGCCCGCT C ATGATCAA	1369
1254	GAGCGUUU CUGAUGAG X CGAA AUCAUGAG	125	CTCATGAT C AAACGCTC	1370
1262	UCUUCUUA CUGAUGAG X CGAA AGCGUUUG	126	CAAACGCT C TAAGAAGA	1371
1264	GUUCUUCU CUGAUGAG X CGAA AGAGCGUU	127	AACGCTCT A AGAAGAAC	1372
1283	UCAGGGAC CUGAUGAG X CGAA AGGCCAGG	128	CCTGGCCT T GTCCCTGA	1373
1286	CCGUCAGG CUGAUGAG X CGAA ACAAGGCC	129	GGCCTTGT C CCTGACGG	1374
1308	AAGGCACU CUGAUGAG X CGAA ACCAUCUG	130	CAGATGGT C AGTGCCTT	1375
1316	CAUCCAAC CUGAUGAG X CGAA AGGCACUG	131	CAGTGCCT T GTTGGATG	1376
1319	CAGCAUCC CUGAUGAG X CGAA ACAAGGCA	132	TGCCITGT T GGATGCTG	1377
1338	GAAUAGAG CUGAUGAG X CGAA AUGGGGGG	133	CCCCCCAT A CTCTATTC	1378
1341	UCGGAAUA CUGAUGAG X CGAA AGUAUGGG	134	CCCATACT C TATTCCGA	1379
1343	ACUCGGAA CUGAUGAG X CGAA AGAGUAUG	135	CATACTCT A TTCCGAGT	1380
1345	AUACUCGG CUGAUGAG X CGAA AUAGAGUA	136	TACTCTAT T CCGAGTAT	1381
1346	CAUACUCG CUGAUGAG X CGAA AAUAGAGU	137	ACTCTATT C CGAGTATG	1382
1352	UAGGAUCA CUGAUGAG X CGAA ACUCGGAA	138	TTCCGAGT A TGATCCTA	1383
1357	UCUGGUAG CUGAUGAG X CGAA AUCAUACU	139	AGTATGAT C CTACCAGA	1384
1360	GGGUCUGG CUGAUGAG X CGAA AGGAUCAU	140	ATGATCCT A CCAGACCC	1385
1370	CUUCACUG CUGAUGAG X CGAA AGGGUCUG	141	CAGACCTT C CAGTGAAG	1386
1371	GCUUCACU CUGAUGAG X CGAA AAGGGUCU	142	AGACCTT C AGTGAAGC	1387
1381	CAUCAUCG CUGAUGAG X CGAA AGCUUCAC	143	GTGAAGCT T CGATGATG	1388
1382	CCAUCAUC CUGAUGAG X CGAA AAGCUUCA	144	TGAAGCTT C GATGATGG	1389
1394	UGGUUCAGU CUGAUGAG X CGAA AGCCCAUC	145	GATGGGCT T ACTGACCA	1390
1395	UUGGUUCAG CUGAUGAG X CGAA AAGCCCAU	146	ATGGGCTT A CTGACCAA	1391
1425	AUCAUGUG CUGAUGAG X CGAA ACCAGCUC	147	GAGCTGGT T CACATGAT	1392
1426	GAUCAUGU CUGAUGAG X CGAA AACCAUCU	148	AGCTGGTT C ACATGATC	1393
1434	GCCCAGUU CUGAUGAG X CGAA AUCAUGUG	149	CACATGAT C AACTGGGC	1394
1460	AAUCCACA CUGAUGAG X CGAA AGCCUGGC	150	GCCAGGCT T TGTGGATT	1395
1461	AAAUCAC CUGAUGAG X CGAA AAGCCUGG	151	CCAGGCTT T GTGGATT	1396

5	1468	GAGGGUCA CUGAUGAG X CGAA AUCCACAA	152	TTGTGGATT TGACCCCTC	1397
	1469	GGAGGGUC CUGAUGAG X CGAA AAUCCACA	153	TGTGGATT T GACCCCTCC	1398
	1476	UGAUCAUG CUGAUGAG X CGAA AGGGUCAA	154	TTGACCCCTC CATGATCA	1399
	1483	GUGGACCU CUGAUGAG X CGAA AUCAUGGA	155	TCCATGATC AGGTCCAC	1400
	1488	AGAAGGGUG CUGAUGAG X CGAA ACCUGAUC	156	GATCAGGT C CACCTTCT	1401
	1494	CAUUCUAG CUGAUGAG X CGAA AGGUGGAC	157	GTCCACCTT CTAGAATG	1402
	1495	ACAUUCUA CUGAUGAG X CGAA AAGGUGGA	158	TCCACCTT C TAGAATGT	1403
10	1497	GCACAUUC CUGAUGAG X CGAA AGAAGGUG	159	CACCTTCTA GAATGTGC	1404
	1512	AGGAUCUC CUGAUGAG X CGAA AGCCAGGC	160	GCCTGGCTA GAGATCCT	1405
	1518	AUCAUCAG CUGAUGAG X CGAA AUCUCUAG	161	CTAGAGATC CTGATGAT	1406
	1527	ACGAGACC CUGAUGAG X CGAA AUCAUCAG	162	CTGATGATT GGTCTCGT	1407
	1531	CCAGACGA CUGAUGAG X CGAA ACCAAUCA	163	TGATTGGTC TCGTCTGG	1408
15	1533	CGCCAGAC CUGAUGAG X CGAA AGACCAAU	164	ATTGGTCTC GTCTGGCG	1409
	1536	GAGCGCCA CUGAUGAG X CGAA ACGAGACC	165	GGTCTCGTC TGGCGCTC	1410
	1544	GCUCCAUG CUGAUGAG X CGAA AGCGCCAG	166	CTGGCGCTC CATGGAGC	1411
	1566	GCAAACAG CUGAUGAG X CGAA AGCUUCAC	167	GTGAAGCTA CTGTTTGC	1412
20	1571	UAGGAGCA CUGAUGAG X CGAA ACAGUAGC	168	GCTACTGT T TGCTCCTA	1413
	1572	UUAGGAGC CUGAUGAG X CGAA AACAGUAG	169	CTACTGTT T GCTCCTAA	1414
	1576	CAAGUUAG CUGAUGAG X CGAA AGCAAACA	170	TGTTTGCTC CTAACTTG	1415
	1579	GAGCAAGU CUGAUGAG X CGAA AGGAGCAA	171	TTGCTCCTA ACTTGCTC	1416
	1583	CCAAGAGC CUGAUGAG X CGAA AGUUAGGA	172	TCCTAACTT GCTCTTGG	1417
	1587	CUGUCCAA CUGAUGAG X CGAA AGCAAGUU	173	AACTTGCTC TTGGACAG	1418
	1589	UCCUGUCC CUGAUGAG X CGAA AGAGCAAG	174	CTTGCTCTT GGACAGGA	1419
	1614	AUGCCCUC CUGAUGAG X CGAA ACACAUUU	175	AAATGTGT A GAGGGCAT	1420
	1632	AUGUCGAA CUGAUGAG X CGAA AUCUCCAC	176	GTGGAGATC TTGACAT	1421
30	1634	GCAUGUCG CUGAUGAG X CGAA AGAUCUCC	177	GGAGATCTT CGACATGC	1422
	1635	AGCAUGUC CUGAUGAG X CGAA AAGAUCUC	178	GAGATCTT C GACATGCT	1423
	1651	AGAUGAUG CUGAUGAG X CGAA AGCCAGCA	179	TGCTGGCTA CATCATCT	1424
	1655	ACCGAGAU CUGAUGAG X CGAA AUGUAGCC	180	GGCTACATC ATCTCGGT	1425
	1658	GGAACCGA CUGAUGAG X CGAA AUGAUGUA	181	TACATCATC TCGGTTCC	1426
	1660	CGCGAACCC CUGAUGAG X CGAA AGAUGAUG	182	CATCATCTC GGTTCCGC	1427

1664	UCAUGCGG CUGAUGAG X CGAA ACCGAGAU	183	ATCTCGGT T CCGCATGA	1428
1665	AUCAUGCG CUGAUGAG X CGAA AACCGAGA	184	TCTCGGTT C CGCATGAT	1429
1678	UCCCUGCA CUGAUGAG X CGAA AUUCAUCA	185	TGATGAAT C TGCAGGGAA	1430
1694	GGCACACCA CUGAUGAG X CGAA ACUCCUCU	186	AGAGGAGTT T GTGTGCC	1431
1695	AGGCACAC CUGAUGAG X CGAA AACUCCUC	187	GAGGGAGTT GTGTGCCT	1432
1704	AUAGAUUU CUGAUGAG X CGAA AGGCACAC	188	GTGTGCCT C AAATCTAT	1433
1709	AAAUAUA CUGAUGAG X CGAA AUUUGAGG	189	CCTCAAAT C TATTATTT	1434
1711	CAAAUAUA CUGAUGAG X CGAA AGAUUUGA	190	TCAAATCT A TTATTTG	1435
1713	AGCAAAAU CUGAUGAG X CGAA AUAGAUU	191	AAATCTAT T ATTTGCT	1436
1714	AAGCAAAA CUGAUGAG X CGAA AAUAGAUU	192	AATCTATT A TTTTGCTT	1437
1716	UUAAGCAA CUGAUGAG X CGAA AUAAUAGA	193	TCTATTAT T TTGCTTAA	1438
1717	AUUAAGCA CUGAUGAG X CGAA AAUAAUAG	194	CTATTATTT TGCTTAAT	1439
1718	AAUUAAGC CUGAUGAG X CGAA AAAUAAUA	195	TATTATTTT GCCTTAATT	1440
1722	CCAGAAAU CUGAUGAG X CGAA AGCAAAAU	196	ATTTGCTT AATTCTGG	1441
1723	UCCAGAAU CUGAUGAG X CGAA AAGCAAAA	197	TTTGCTT A ATTCTGGA	1442
1726	CACUCCAG CUGAUGAG X CGAA AUUAAGCA	198	TGCTTAATT CTGGAGTG	1443
1727	ACACUCCA CUGAUGAG X CGAA AAUUAAGC	199	GCTTAATT C TGGAAGTGT	1444
1736	AAAUGUG CUGAUGAG X CGAA ACACUCCA	200	TGGAGTGT A CACATITC	1445
1742	UGGACAGA CUGAUGAG X CGAA AUGUGUAC	201	GTACACATT TCTGTCCA	1446
1743	CUGGACAG CUGAUGAG X CGAA AAUGUGUA	202	TACACATT T CTGTCCAG	1447
1744	GCUGGGACA CUGAUGAG X CGAA AAAUGUGU	203	ACACATT T GTGCCAGC	1448
1748	GGGUGGCUG CUGAUGAG X CGAA ACAGAAAU	204	ATTTCTGT C CAGCACCC	1449
1763	CUUCCAGA CUGAUGAG X CGAA ACUUCAGG	205	CCTGAAGT C TCTGGAAG	1450
1765	CUCUUCCA CUGAUGAG X CGAA AGACUUCA	206	TGAAGTCT C TGGAAGAG	1451
1783	UCGGUGGA CUGAUGAG X CGAA AUGGUCCU	207	AGGACCATA TCCACCGA	1452
1785	ACUCGGUG CUGAUGAG X CGAA AUAUGGUC	208	GACCATAT C CACCGAGT	1453
1794	UUGUCCAG CUGAUGAG X CGAA ACUCCGUG	209	CACCGAGT C CTGGACAA	1454
1806	GUGUCUGU CUGAUGAG X CGAA AUCUUGUC	210	GACAAGAT C ACAGACAC	1455
1816	GUGGAUCA CUGAUGAG X CGAA AGUGUCUG	211	CAGACACT T TGATCCAC	1456
1817	GGUGGAUC CUGAUGAG X CGAA AAGUGUCU	212	AGACACIT T GATCCACC	1457
1821	AUCAGGUG CUGAUGAG X CGAA AUCAAAGU	213	ACTTTGAT C CACCTGAT	1458

1881	AUGAGGGAG CUGAUGAG X CGAA AGCUUGGC	214	GCCCCAGCT C CTCCTCAT	1459
1884	AGGAUGAG CUGAUGAG X CGAA AGGAGCUG	215	CAGCTCCT C CTCATCCT	1460
1887	GAGAGGAU CUGAUGAG X CGAA AGGAGGAG	216	CTCCTCCT C ATCCTCTC	1461
1890	UGGGAGAG CUGAUGAG X CGAA AUGAGGAG	217	CTCCTCAT C CTCTCCCA	1462
1893	AUGUGGGAG CUGAUGAG X CGAA AGGAUGAG	218	CTCATCCT C TCCCACAT	1463
1895	UGAUGUGG CUGAUGAG X CGAA AGAGGAUG	219	CATCCTCT C CCACATCA	1464
1902	AUGUGCCU CUGAUGAG X CGAA AUGUGGGA	220	TCCCACAT C AGGCACAT	1465
1915	GCCUUUGU CUGAUGAG X CGAA ACUCAUGU	221	ACATGAGT A ACAAAGGC	1466
1933	GCUGUACA CUGAUGAG X CGAA AUGCUCCA	222	TGGAGCAT C TGTACAGC	1467
1937	UCAUGGUG CUGAUGAG X CGAA ACAGAUGC	223	GCATCTGT A CAGCATGA	1468
1968	AGGUCAUA CUGAUGAG X CGAA AGGGGCAC	224	GTGCCCCT C TATGACCT	1469
1970	GCAGGUCA CUGAUGAG X CGAA AGAGGGGC	225	GCCCCTCT A TGACCTGC	1470
2007	GGCGCAUG CUGAUGAG X CGAA AGGCGGUG	226	CACCGCCT A CATGCGCC	1471
2020	UCCACGGC CUGAUGAG X CGAA AGUGGGCG	227	CGCCCACT A GCCGTGGA	1472
2036	CCUCCACG CUGAUGAG X CGAA AUGCCCCU	228	AGGGGCAT C CGTGGAGG	1473
2063	CAGUGGCC CUGAUGAG X CGAA AGUGGCCU	229	AAGCCACTT GGCCACTG	1474
2078	AUGAAGUA CUGAUGAG X CGAA AGCCCGCA	230	TGCGGGCT C TACTTCAT	1475
2080	CGAUGAAG CUGAUGAG X CGAA AGAGCCCC	231	CGGGCTCT A CTTCATCG	1476
2083	AUGCGAUG CUGAUGAG X CGAA AGUAGAGC	232	GCTCTACTT CATCGCAT	1477
2084	AAUGCGAU CUGAUGAG X CGAA AAGUAGAG	233	CTCTACTT C ATCGCATT	1478
2087	AGGAAUGC CUGAUGAG X CGAA AUGAAGUA	234	TACTTCAT C GCATTCT	1479
2092	UUGCAAGG CUGAUGAG X CGAA AUGCGAUG	235	CATCGCATT CCTTGCAA	1480
2093	UUUGCAAG CUGAUGAG X CGAA AAUGCGAU	236	ATCGCATT C TTGCAAA	1481
2096	ACUUUUGC CUGAUGAG X CGAA AGGAAUGC	237	GCATTCTT GCAAAAGT	1482
2105	UGAUGUAA CUGAUGAG X CGAA ACUUUUGC	238	GCAAAAGT A TTACATCA	1483
2107	CGUGAUGU CUGAUGAG X CGAA AUACUUUU	239	AAAAGTATT ACATCACG	1484
2108	CCGUGAUG CUGAUGAG X CGAA AAUACUUU	240	AAAGTATT A CATCACGG	1485
2112	UCCCCCGU CUGAUGAG X CGAA AUGUAAUA	241	TATTACAT C ACGGGGGA	1486
2131	GGCAGGGAG CUGAUGAG X CGAA ACCCUCUG	242	CAGAGGGTT TCCCTGCC	1487
2132	UGGCAGGG CUGAUGAG X CGAA AACCCUCU	243	AGAGGGTTT CCCTGCCA	1488
2133	GUGGCAGG CUGAUGAG X CGAA AAACCCUC	244	GAGGGTTT CCTGCCAC	1489

2145	AGCUCUCA CUGAUGAG X CGAA ACUGUGGC	245	GCCACAGT C TGAGAGCT	1490
2154	GAGCCAGG CUGAUGAG X CGAA AGCUCUCA	246	TGAGAGCT C CCTGGCTC	1491
2162	CCGUGUGG CUGAUGAG X CGAA AGCCAGGG	247	CCCTGGCT C CCACACGG	1492
2172	AUUAUCUG CUGAUGAG X CGAA ACCGUGUG	248	CACACGGT T CAGATAAT	1493
2173	GAUUAUCU CUGAUGAG X CGAA AACCGUGU	249	ACACGGTT C AGATAATC	1494
2178	GCAGGGAU CUGAUGAG X CGAA AUCUGAAC	250	GTTAGAT A ATCCCTGC	1495
2181	GCAGCAGG CUGAUGAG X CGAA AUUAUCUG	251	CAGATAAT C CCTGCTGC	1496
2192	GAGGGUA CUGAUGAG X CGAA AUGCAGCA	252	TGCTGCAT T TTACCCCTC	1497
2193	UGAGGGUA CUGAUGAG X CGAA AAUGCAGC	253	GCTGCATT T TACCCCTCA	1498
2194	AUGAGGGU CUGAUGAG X CGAA AAAUGCAG	254	CTGCATTT T ACCCTCAT	1499
2195	GAUGAGGG CUGAUGAG X CGAA AAAAUGCA	255	TGCATTTT A CCCTCATC	1500
2200	UGCAUGAU CUGAUGAG X CGAA AGGGUAAA	256	TTTACCCCT C ATCATGCA	1501
2203	UGGUGCAU CUGAUGAG X CGAA AUGAGGGU	257	ACCCCTCAT C ATGCACCA	1502
2214	UUUGGCUA CUGAUGAG X CGAA AGUGGUGC	258	GCACCACT T TAGCCAAA	1503
2215	AUUUGGCU CUGAUGAG X CGAA AAGUGGUG	259	CACCACTT T AGCCAAAT	1504
2216	AAUUUGGC CUGAUGAG X CGAA AAAGUGGU	260	ACCACTTT A GCCAAATT	1505
2224	GGAGACAG CUGAUGAG X CGAA AUUUGGCU	261	AGCCAAATT T CTGTCCTCC	1506
2225	AGGAGACA CUGAUGAG X CGAA AAUUUGGC	262	GCCAAATT C TGTCTCCT	1507
2229	AUGCAGGA CUGAUGAG X CGAA ACAGAAUU	263	AATTCTGT C TCCTGCAT	1508
2231	GUAUGCAG CUGAUGAG X CGAA AGACAGAA	264	TTCTGTCT C CTGCATAC	1509
2238	CCGGAGUG CUGAUGAG X CGAA AUGCAGGA	265	TCCTGCAT A CACTCCGG	1510
2243	GCAUGCCG CUGAUGAG X CGAA AGUGUAUG	266	CATACACT C CGGCATGC	1511
2254	UGGUGUUG CUGAUGAG X CGAA AUGCAUGC	267	GCATGCAT C CAACACCA	1512
2269	CAUCUAGA CUGAUGAG X CGAA AGCCAUG	268	CAATGGCT T TCTAGATG	1513
2270	UCAUCUAG CUGAUGAG X CGAA AAGCCAUU	269	AATGGCTT T CTAGATGA	1514
2271	CUCAUCUA CUGAUGAG X CGAA AAAGCCAU	270	ATGGCTTT C TAGATGAG	1515
2273	CACUCAUC CUGAUGAG X CGAA AGAAAGCC	271	GGCTTTCT A GATGAGTG	1516
2287	AGCAAAUG CUGAUGAG X CGAA AUGGCCAC	272	GTGGCCATT C ATTTGCT	1517
2288	AAGCAAAU CUGAUGAG X CGAA AAUGGCCA	273	TGGCCATT C ATTTGCTT	1518
2291	AGCAAGCA CUGAUGAG X CGAA AUGAAUGG	274	CCATTCACT T TGCTTGCT	1519
2292	GAGCAAGC CUGAUGAG X CGAA AAUGAAUG	275	CATTCACTT GCTTGCTC	1520

2296	AACUGAGC CUGAUGAG X CGAA AGCAAAUG	276	CATTTGCT T GCTCAGTT	1521
2300	UAAGAACU CUGAUGAG X CGAA AGCAAGCA	277	TGCTTGCT C AGTTCTTA	1522
2304	CCACUAAG CUGAUGAG X CGAA ACUGAGCA	278	TGCTCAGT T CTTAGTGG	1523
2305	GCCACUAA CUGAUGAG X CGAA AACUGAGC	279	GCTCAGTT C TTAGTGGC	1524
2307	GUGCCACU CUGAUGAG X CGAA AGAACUGA	280	TCAGTTCT T AGTGGCAC	1525
2308	UGUGCCAC CUGAUGAG X CGAA AAGAACUG	281	CAGTTCTT A GTGGCACA	1526
2318	AGACAGAA CUGAUGAG X CGAA AUGUGCCA	282	TGGCACAT C TTCTGTCT	1527
2320	GAAGACAG CUGAUGAG X CGAA AGAUGUGC	283	GCACATCT T CTGTCTTC	1528
2321	AGAAGACA CUGAUGAG X CGAA AAGAUGUG	284	CACATCTT C TGTCTTCT	1529
2325	CAACAGAA CUGAUGAG X CGAA ACAGAAGA	285	TCTTCTGT C TTCTGTIG	1530
2327	CCCAACAG CUGAUGAG X CGAA AGACAGAA	286	TTCTGTCT T CTGTTGGG	1531
2328	UCCCAACA CUGAUGAG X CGAA AAGACAGA	287	TCTGTCTT C TGTTGGGA	1532
2332	CUGUUCCC CUGAUGAG X CGAA ACAGAAGA	288	TCTTCTGT T GGGAACAG	1533
2351	AGCCUUGG CUGAUGAG X CGAA AUCCCUU	289	AAAGGGATT CCAAGGCT	1534
2352	UAGCCUUG CUGAUGAG X CGAA AAUCCUU	290	AAGGGATT C CAAGGCTA	1535
2360	CAAAGAUU CUGAUGAG X CGAA AGCCUUGG	291	CCAAGGCT A AATCTTTG	1536
2364	GUUACAAA CUGAUGAG X CGAA AUUUAGCC	292	GGCTAAAT C TTTGTAAC	1537
2366	CUGUUACA CUGAUGAG X CGAA AGAUUUAG	293	CTAAATCT T TGTAACAG	1538
2367	GCUGUUAC CUGAUGAG X CGAA AAGAUUUA	294	TAAATCTT T GTAACAGC	1539
2370	AGAGCUGU CUGAUGAG X CGAA ACAAAGAU	295	ATCTTTGT A ACAGCTCT	1540
2377	GGGAAAGA CUGAUGAG X CGAA AGCUGUUA	296	TAACAGCT C TCTTCCCC	1541
2379	GGGGGAAA CUGAUGAG X CGAA AGAGCUGU	297	ACAGCTCT C TTTCCCCC	1542
2381	AAGGGGGA CUGAUGAG X CGAA AGAGAGCU	298	AGCTCTCT T TCCCCCTT	1543
2382	CAAGGGGG CUGAUGAG X CGAA AAGAGAGC	299	GCTCTCTT T CCCCCCTT	1544
2383	GCAAGGGG CUGAUGAG X CGAA AAAGAGAG	300	CTCTCTTT C CCCCTTGC	1545
2389	AACAUAGC CUGAUGAG X CGAA AGGGGGAA	301	TTCCCCCT T GCTATGTT	1546
2393	UAGUAACA CUGAUGAG X CGAA AGCAAGGG	302	CCCTTGCT A TGTTACTA	1547
2397	CGCUUAGU CUGAUGAG X CGAA ACAUAGCA	303	TGCTATGTT ACTAAGCG	1548
2398	ACGCUUAG CUGAUGAG X CGAA AACAUAGC	304	GCTATGTT A CTAAGCGT	1549
2401	CUCACGCU CUGAUGAG X CGAA AGUAACAU	305	ATGTTACT A AGCGTGAG	1550
2413	GCUACGGG CUGAUGAG X CGAA AUCCUCAC	306	GTGAGGATT CCCGTAGC	1551

2414	AGCUACGG CUGAUGAG X CGAA AAUCCUCA	307	TGAGGATT C CCGTAGCT	1552
2419	UGAAGAGC CUGAUGAG X CGAA ACCGGAAU	308	ATTCCCGT A GCTCTTCA	1553
2423	GCUGUGAA CUGAUGAG X CGAA AGCUACGG	309	CCGTAGCT C TTCACAGC	1554
2425	CAGCUGUG CUGAUGAG X CGAA AGAGCUAC	310	GTAGCTCT T CACAGCTG	1555
5	UCAGCUGU CUGAUGAG X CGAA AAGAGCUA	311	TAGCTCTT C ACAGCTGA	1556
2426	CAUAGACU CUGAUGAG X CGAA AGUUCAGC	312	GCTGAACT C AGTCTATG	1557
2438	AACCCAUA CUGAUGAG X CGAA ACUGAGUU	313	AACTCAGT C TATGGGTT	1558
2442	CCAACCCA CUGAUGAG X CGAA AGACUGAG	314	CTCAGTCT A TGGGTTGG	1559
10	UGAGCCCC CUGAUGAG X CGAA ACCCAUAG	315	CTATGGGT T GGGGCTCA	1560
2450	AGUUAUCU CUGAUGAG X CGAA AGCCCCAA	316	TTGGGGCT C AGATAACT	1561
2457	CACAGAGU CUGAUGAG X CGAA AUCUGAGC	317	GCTCAGAT A ACTCTGTG	1562
2462	AAUGCACA CUGAUGAG X CGAA AGUUAUCU	318	AGATAACT C TGTGCATT	1563
15	GUAGCUUA CUGAUGAG X CGAA AUGCACAG	319	CTGTGCAT T TAAGCTAC	1564
2474	AGUAGCUU CUGAUGAG X CGAA AAUGCACA	320	TGTGCATT T AAGCTACT	1565
2475	AAGUAGCU CUGAUGAG X CGAA AAAUGCAC	321	GTGCATTT A AGCTACTT	1566
2476	UCUACAAG CUGAUGAG X CGAA AGCUUAAA	322	TTTAAGCT A CTTGTAGA	1567
20	GUCUCUAC CUGAUGAG X CGAA AGUAGCUU	323	AAGCTACT T GTAGAGAC	1568
2484	UGGGUCUC CUGAUGAG X CGAA ACAAGUAG	324	CTACTTGT A GAGACCCA	1569
2487	AAAAUGUC CUGAUGAG X CGAA ACUCUCCA	325	TGGAGAGT A GACATTTT	1570
2508	AGAGGCAA CUGAUGAG X CGAA AUGCUAC	326	GTAGACAT T TTGCCTCT	1571
2514	CAGAGGCA CUGAUGAG X CGAA AAUGCUA	327	TAGACATT T TGCCCTTG	1572
2515	UCAGAGGC CUGAUGAG X CGAA AAAUGUCU	328	AGACATT T GCCTCTGA	1573
2516	GUUAUCA CUGAUGAG X CGAA AGGCAAAA	329	TTTGCCCT C TGATAAGC	1574
2521	AAAGUGCU CUGAUGAG X CGAA AUCAGAGG	330	CCTCTGAT A AGCACTT	1575
2526	CAUUUAAA CUGAUGAG X CGAA AGUGCUUA	331	TAAGCACT T TTTAAATG	1576
30	CCAUUUAA CUGAUGAG X CGAA AAGUGCUU	332	AAGCACTT T TTAAATGG	1577
2533	GCCAUUU CUGAUGAG X CGAA AAAGUGCU	333	AGCACTTT T TAAATGGC	1578
2534	AGCCAUUU CUGAUGAG X CGAA AAAAGUGC	334	GCACTTT T AAATGGCT	1579
2536	GAGCCAUU CUGAUGAG X CGAA AAAAGUG	335	CACTTTT A AATGGCTC	1580
2537	UAUUCUUA CUGAUGAG X CGAA AGCCAUU	336	AAATGGCT C TAAGAATA	1581
2545	CUUAUUCU CUGAUGAG X CGAA AGAGCCAU	337	ATGGCTCT A AGAATAAG	1582

5	2553	CUGUGGCU CUGAUGAG X CGAA AUUCUUAG	338	CTAAGAAT A AGCCACAG	1583
	2570	CCACUUUA CUGAUGAG X CGAA AUUCUUUG	339	CAAAGAATT TAAAGTGG	1584
	2571	GCCACUUU CUGAUGAG X CGAA AAUUCUUU	340	AAAGAATT T AAAGTGGC	1585
	2572	AGCCACUU CUGAUGAG X CGAA AAAUUCUU	341	AAGAATT T AAGTGGCT	1586
10	2581	AAUUAAG CUGAUGAG X CGAA AGCCACUU	342	AAGTGGCT C CTITAATT	1587
	2584	ACCAAUUA CUGAUGAG X CGAA AGGAGCCA	343	TGGCTCCT T TAATTGGT	1588
	2585	CACCAAUU CUGAUGAG X CGAA AAGGAGCC	344	GGCTCCTT T AATTGGTG	1589
	2586	UCACCAAU CUGAUGAG X CGAA AAAGGAGC	345	GCTCCTT A ATTGGTGA	1590
15	2589	AAGUCACC CUGAUGAG X CGAA AUUAAAGG	346	CCTTAATT GGTGACTT	1591
	2597	CUUUCUCC CUGAUGAG X CGAA AGUCACCA	347	TGGTGAATT GGAGAAAG	1592
	2608	CCUUGACC CUGAUGAG X CGAA AGCUUUCU	348	AGAAAGCT A GGTCAAGG	1593
	2612	AAACCCUU CUGAUGAG X CGAA ACCUAGCU	349	AGCTAGGT C AAGGGTTT	1594
20	2619	CUAUAAUA CUGAUGAG X CGAA ACCCUUGA	350	TCAAGGGTT TATTATAG	1595
	2620	GCUAUAAA CUGAUGAG X CGAA AACCCUUG	351	CAAGGGTT T ATTATAGC	1596
	2621	UGCUAUAA CUGAUGAG X CGAA AAACCCUU	352	AAGGGTTT A TTATAGCA	1597
	2623	GGUGCUAU CUGAUGAG X CGAA AUAAACCC	353	GGGTTTATT ATAGCACC	1598
25	2624	GGGUGCUA CUGAUGAG X CGAA AAUAAAACC	354	GGTTTATT A TAGCACCC	1599
	2626	GAGGGUGC CUGAUGAG X CGAA AUAAUAAA	355	TTTATTAT A GCACCCCTC	1600
	2634	GAAUACAA CUGAUGAG X CGAA AGGGUGCU	356	ACGACCCCT C TTGTATTTC	1601
	2636	AGGAAUAC CUGAUGAG X CGAA AGAGGGUG	357	CACCCCTCT T GTATTCCCT	1602
30	2639	CAUAGGAA CUGAUGAG X CGAA ACAAGAGG	358	CCTCTTGT A TTCTATG	1603
	2641	GCCAUAGG CUGAUGAG X CGAA AUACAAGA	359	TCTTGTATT CCTATGGC	1604
	2642	UGCCAUAG CUGAUGAG X CGAA AAUACAAG	360	CTTGTATT C CTATGGCA	1605
	2645	CAUUGCCA CUGAUGAG X CGAA AGGAAUAC	361	GTATTCCCT A TGGCAATG	1606
	2657	CAUAAAAG CUGAUGAG X CGAA AUGCAUUG	362	CAATGCAT C CTTTTATG	1607
	2660	UUUCAUAA CUGAUGAG X CGAA AGGAUGCA	363	TGCATCCT T TTATGAAA	1608
	2661	CUUUCUA CUGAUGAG X CGAA AAGGAUGC	364	GCATCCCT T TATGAAAG	1609
	2662	ACUUUCAU CUGAUGAG X CGAA AAAGGAUG	365	CATCCTTT T ATGAAAGT	1610
	2663	CACUUUCA CUGAUGAG X CGAA AAAAGGAU	366	ATCCTTTT A TGAAAGTG	1611
	2674	UUAAGGUG CUGAUGAG X CGAA ACCACUUU	367	AAAGTGGT A CACCTTAA	1612
	2680	AAAGCUUU CUGAUGAG X CGAA AGGUGUAC	368	GTACACCT T AAAGCTTT	1613

2681	AAAAGCUU CUGAUGAG X CGAA AAGGUGUA	369	TACACCTT A AAGCTTTT	1614
2687	UCAUUAUA CUGAUGAG X CGAA AGCUUUAA	370	TTAAAGCT T TTATATGA	1615
2688	GUCAUAUA CUGAUGAG X CGAA AAGCUUUA	371	TAAAGCTT T TATATGAC	1616
2689	AGUCAUUA CUGAUGAG X CGAA AAAGCUUU	372	AAAGCTTT T ATATGACT	1617
2690	CAGUCAUA CUGAUGAG X CGAA AAAAGCUU	373	AAGCTTTT A TATGACTG	1618
2692	UACAGUCA CUGAUGAG X CGAA AUAAAAGC	374	GCTTTTAT A TGACTGTA	1619
2700	UACUCUGC CUGAUGAG X CGAA ACAGUCAU	375	ATGACTGT A GCAGAGTA	1620
2708	UCACCAGA CUGAUGAG X CGAA ACUCUGCU	376	AGCAGAGT A TCTGGTGA	1621
2710	AAUCACCA CUGAUGAG X CGAA AUACUCUG	377	CAGAGTAT C TGGTGATT	1622
2718	GAAUUGAC CUGAUGAG X CGAA AUCACCAG	378	CTGGTGAT T GTCAATTTC	1623
2721	AGUQAAUU CUGAUGAG X CGAA ACAAUCAC	379	GTGATTGT C AATTCACT	1624
2725	GGGAAGUG CUGAUGAG X CGAA AUUGACAA	380	TTGTCAATT CACITCCCC	1625
2726	GGGGAAGU CUGAUGAG X CGAA AAUUGACA	381	TGTCAATT C ACTTCCCC	1626
2730	AUAGGGGG CUGAUGAG X CGAA AGUGAAU	382	AATTCACT T CCCCTAT	1627
2731	UAUAGGGG CUGAUGAG X CGAA AAGUGAAU	383	ATTCACTT C CCCCTATA	1628
2737	UAUUCCUA CUGAUGAG X CGAA AGGGGGAA	384	TTCCCCCT A TAGGAATA	1629
2739	UGUAUUCC CUGAUGAG X CGAA AUAGGGGG	385	CCCCCTAT A GGAATACA	1630
2745	GCCCCUUG CUGAUGAG X CGAA AUUCCUAU	386	ATAGGAAT A CAAGGGGC	1631
2772	AACUAGGG CUGAUGAG X CGAA AUCUGCCU	387	AGGCAGAT C CCCTAGTT	1632
2777	UGGCCAAC CUGAUGAG X CGAA AGGGGAUC	388	GATCCCCCT A GTGGCCA	1633
2780	UCUUGGCC CUGAUGAG X CGAA ACUAGGGG	389	CCCCTAGT T GGCCAAGA	1634
2791	GUUAAAUA CUGAUGAG X CGAA AGUCUUGG	390	CCAAGACTT ATTAAAC	1635
2792	AGUUAAAA CUGAUGAG X CGAA AAGUCUUG	391	CAAGACTT A TTTAACT	1636
2794	CAAGUUAA CUGAUGAG X CGAA AUAAGUCU	392	AGACTTATT T TAACTTG	1637
2795	UCAAGUUA CUGAUGAG X CGAA AAUAAGUC	393	GACTTATT T TAACTTGA	1638
2796	AUCAAGUU CUGAUGAG X CGAA AAAUAAGU	394	ACTTATTIT T AACTTGAT	1639
2797	UAUCAAGU CUGAUGAG X CGAA AAAUAAG	395	CITTTTTT A ACTTGATA	1640
2801	AGUGUAUC CUGAUGAG X CGAA AGUAAAAA	396	TTTTAACT T GATACACT	1641
2805	CUGCAGUG CUGAUGAG X CGAA AUCAAGUU	397	AACTTGAT A CACTGCAG	1642
2816	ACACUCUG CUGAUGAG X CGAA AUCUGCAG	398	CTGCAGATT CAGAGTGT	1643
2817	GACACUCU CUGAUGAG X CGAA AAUCUGCA	399	TGCAGATT C AGAGTGTC	1644

5	2825	AGCUUCAG CUGAUGAG X CGAA ACACUCUG	400	CAGAGTGT C CTGAAGCT	1645
	2834	CAGAGGCA CUGAUGAG X CGAA AGCUUCAG	401	CTGAAGCT C TGCCTCTG	1646
	2840	GAAAGCCA CUGAUGAG X CGAA AGGCAGAG	402	CTCTGCCT C TGGCTTTC	1647
	2846	UGACCGGA CUGAUGAG X CGAA AGCCAGAG	403	CTCTGGCT T TCCGGTCA	1648
10	2847	AUGACCGG CUGAUGAG X CGAA AAGCCAGA	404	TCTGGCTT T CCGGTCA	1649
	2848	CAUGACCG CUGAUGAG X CGAA AAAGCCAG	405	CTGGCTT C CGGTCA	1650
	2853	GAACCCAU CUGAUGAG X CGAA ACCGGAAA	406	TTTCCGGT C ATGGGTT	1651
	2860	UUAAACUGG CUGAUGAG X CGAA ACCCAUGA	407	TCATGGGT T CCAGTTAA	1652
15	2861	AUUAACUG CUGAUGAG X CGAA AACCCAUG	408	CATGGGTT C CAGTTAAT	1653
	2866	CAUGAAUU CUGAUGAG X CGAA ACUGGAAC	409	GTTCAGTT AATTCA	1654
	2867	GCAUGAAU CUGAUGAG X CGAA AACUGGAA	410	TTCCAGTT A ATTCA	1655
	2870	GAGGCAUG CUGAUGAG X CGAA AUUAAACUG	411	CAGTTAAT T CATGCCTC	1656
20	2871	GGAGGCAU CUGAUGAG X CGAA AAUUAACU	412	AGTTAATT C ATGCCTCC	1657
	2878	GUCCAUGG CUGAUGAG X CGAA AGGCAUGA	413	TCATGCCT C CCATGGAC	1658
	2889	GCUCUCCA CUGAUGAG X CGAA AGGUCCAU	414	ATGGACCT A TGGAGAGC	1659
	2905	CUAAGAUC CUGAUGAG X CGAA ACUUGUUG	415	CAACAAGTT GATCTTAG	1660
25	2909	UUAACUAA CUGAUGAG X CGAA AUCAACUU	416	AAGTTGAT C TTAGTTAA	1661
	2911	ACUUAACU CUGAUGAG X CGAA AGAUCAAC	417	GTGATCT T AGTTAAGT	1662
	2912	GACUUAAC CUGAUGAG X CGAA AAGAUCAA	418	TTGATCTT A GTTAAGTC	1663
	2915	GGAGACUU CUGAUGAG X CGAA ACUAAGAU	419	ATCTTAGT T AAGTCTCC	1664
30	2916	GGGAGACU CUGAUGAG X CGAA AACUAAGA	420	TCTTAGTT A AGTCTCCC	1665
	2920	UAUAGGGA CUGAUGAG X CGAA ACUUAACU	421	AGTTAAGT C TCCCTATA	1666
	2922	CAUAUAGG CUGAUGAG X CGAA AGACUUA	422	TTAAGTCT C CCTATATG	1667
	2926	CCCUCAUA CUGAUGAG X CGAA AGGGAGAC	423	GTCTCCCT A TATGAGGG	1668
	2928	AUCCCUCA CUGAUGAG X CGAA AUAGGGAG	424	CTCCCTAT A TGAGGGAT	1669
	2937	CAGGAACU CUGAUGAG X CGAA AUCCCUCA	425	TGAGGGAT A AGTCCCTG	1670
	2941	AAAUCAGG CUGAUGAG X CGAA ACUUAUCC	426	GGATAAGT T CCTGATT	1671
	2942	AAAUCAG CUGAUGAG X CGAA AACUUAUC	427	GATAAGTT C CTGATT	1672
	2948	AAAACAAA CUGAUGAG X CGAA AUCAGGA	428	TTCCCTGAT T TTGTTT	1673
	2949	AAAACAA CUGAUGAG X CGAA AAUCAGGA	429	TCCTGATT T TTGTTT	1674
	2950	AAAAAACAA CUGAUGAG X CGAA AAAUCAGG	430	CCTGATT T TGTTT	1675

5	2951	AUAAAAC CUGAUGAG X CGAA AAAUCAG	431	CTGATTTT T GTTTTTAT	1676
	2954	AAAAUAAA CUGAUGAG X CGAA ACAAAAAU	432	ATTTTTGT T TTTATTTT	1677
	2955	AAAAAUAA CUGAUGAG X CGAA AACAAAAA	433	TTTTGTT T TTATTTT	1678
	2956	CAAAAAUA CUGAUGAG X CGAA AAACAAAA	434	TTTGTTTT T TTTTTTG	1679
	2957	ACAAAAAU CUGAUGAG X CGAA AAAACAAA	435	TTTGTTTT T ATTTTTGT	1680
	2958	CACAAAAA CUGAUGAG X CGAA AAAAACAA	436	TTGTTTTT A TTTTTGTG	1681
10	2960	AACACAAA CUGAUGAG X CGAA AUAAAAC	437	GTTTTATT T TTTGTGTT	1682
	2961	UAACACAA CUGAUGAG X CGAA AAUAAAAA	438	TTTTTATT T TTGTGITA	1683
	2962	GUAACACA CUGAUGAG X CGAA AAAUAAAA	439	TTTATTTT T TTGTGTTAC	1684
	2963	UGUAACAC CUGAUGAG X CGAA AAAUAAA	440	TTTATTTT T GTGTTACA	1685
	2968	UCUUUUGU CUGAUGAG X CGAA ACACAAAA	441	TTTGTGTT ACAGAAAGA	1686
	2969	UUCUUUUG CUGAUGAG X CGAA AACACAAA	442	TTTGTGTT A CAAAGAGA	1687
15	2984	CAGGGAGG CUGAUGAG X CGAA AGGGCUUU	443	AAAGCCCT C CCTCCCTG	1688
	2988	AGUUCAGG CUGAUGAG X CGAA AGGGAGGG	444	CCCTCCCT C CCTGAAC	1689
	2997	CUUACUGC CUGAUGAG X CGAA AGUUCAGG	445	CCTGAAC T GCAGTAAG	1690
	3003	GCUGACCU CUGAUGAG X CGAA ACUGCAAG	446	CTTGCAGT A AGGTCAGC	1691
20	3008	CUGAAGCU CUGAUGAG X CGAA ACCUUACU	447	AGTAAGGT C AGCTTCAG	1692
	3013	AGGUCCUG CUGAUGAG X CGAA AGCUGACC	448	GGTCAGCT T CAGGACCT	1693
	3014	CAGGUCCU CUGAUGAG X CGAA AAGCUGAC	449	GTCAGCTT C AGGACCTG	1694
	3024	CCACACUG CUGAUGAG X CGAA ACAGGUCC	450	GGACCTGT T CCAGTGGG	1695
	3025	GCCCACUG CUGAUGAG X CGAA AACAGGUC	451	GACCTGTT C CAGTGGGC	1696
25	3039	GAUCCAAG CUGAUGAG X CGAA ACAGUGCC	452	GGCACTGT A CTTGGATC	1697
	3042	GAAGAUCC CUGAUGAG X CGAA AGUACAGU	453	ACTGTACT T GGATCTTC	1698
	3047	GCCGGGAA CUGAUGAG X CGAA AUCCAAGU	454	ACTTGGAT C TTCCCGGC	1699
	3049	ACGCCGGG CUGAUGAG X CGAA AGAUCCAA	455	TTGGATCT T CCCGGCGT	1700
30	3050	CACGCCGG CUGAUGAG X CGAA AAGAUCCA	456	TGGATCTT C CCGCGTGT	1701
	3068	CCCUGUGU CUGAUGAG X CGAA AGGCACAC	457	GTGTGCCTT ACACAGGG	1702
	3069	CCCCUGUG CUGAUGAG X CGAA AAGGCACA	458	TGTGCCTT A CACAGGGG	1703
	3086	CCACAGUG CUGAUGAG X CGAA ACAGUUCA	459	TGAACGTGTT CACTGTGG	1704
	3087	ACCACAGU CUGAUGAG X CGAA AACAGUUC	460	GAACGTGTT C ACTGTGGT	1705
	3112	CUACCAUU CUGAUGAG X CGAA ACCCUAU	461	ATGAGGGT A AATGGTAG	1706

3119	CUUUCAAC CUGAUGAG X CGAA ACCAUUUA	462	TAAATGGT A GTTGAAG	1707
3122	CUCCUUUC CUGAUGAG X CGAA ACUACCAU	463	ATGGTAGT T GAAAGGAG	1708
3146	CUAAAUGC CUGAUGAG X CGAA ACACCAGG	464	CCTGGTGT T GCATTTAG	1709
3151	CAGGGCUA CUGAUGAG X CGAA AUGCAACA	465	TGTTGCATT TAGCCCTG	1710
3152	CCAGGGCU CUGAUGAG X CGAA AAUGCAAC	466	GTTGCATT TAGCCCTGG	1711
3153	CCCAGGGC CUGAUGAG X CGAA AAAUGCAA	467	TTGCATT A GCCCTGGG	1712
3179	UGCACAAG CUGAUGAG X CGAA ACUGUUCA	468	TGAACAGT A CTTGTGCA	1713
3182	UCCUGCAC CUGAUGAG X CGAA AGUACUGU	469	ACAGTACT T GTGCAGGA	1714
3192	GCCACAAC CUGAUGAG X CGAA AUCCUGCA	470	TGCAGGATT GTTGTGGC	1715
3195	GUAGCCAC CUGAUGAG X CGAA ACAAUCCU	471	AGGATTGT T GTGGCTAC	1716
3202	UUCUCUAG CUGAUGAG X CGAA AGCCACAA	472	TTGTGGCT A CTAGAGAA	1717
3205	UUGUUUCUC CUGAUGAG X CGAA AGUAGCCA	473	TGGCTACT A GAGAACAA	1718
3224	UUCUGCCC CUGAUGAG X CGAA ACUUUCC	474	GGGAAAGT A GGGCAGAA	1719
3240	CAGAACUG CUGAUGAG X CGAA AUCCAGUU	475	AACTGGAT A CAGTTCTG	1720
3245	GUGCUAG CUGAUGAG X CGAA ACUGUAUC	476	GATACAGT T CTGAGCAC	1721
3246	UGUGCUCU CUGAUGAG X CGAA AACUGUAU	477	ATACAGTT C TGAGCACA	1722
3263	ACCUGAGC CUGAUGAG X CGAA AGUCUGGC	478	GCCAGACT T GCTCAGGT	1723
3267	GGCCACCU CUGAUGAG X CGAA AGCAAGUC	479	GACTTGCT C AGGTGGCC	1724
3293	UUCCUAGG CUGAUGAG X CGAA AGCUGCAG	480	CTGCAGCT A CCTAGGAA	1725
3297	AAUGUUCC CUGAUGAG X CGAA AGGUAGCU	481	AGCTACCT A GGAACATT	1726
3305	CUGCAAGG CUGAUGAG X CGAA AUGUUCC	482	AGGAACAT T CCTTGCAG	1727
3306	UCUGCAAG CUGAUGAG X CGAA AAUGUUCC	483	GGAACATT C TTGCAGA	1728
3309	GGGUCUGC CUGAUGAG X CGAA AGGAAUGU	484	ACATTCCCT GCAGACCC	1729
3323	CCAAAGGC CUGAUGAG X CGAA AUGGGGG	485	CCCCGCATT GCCTTTGG	1730
3328	CACCCCCA CUGAUGAG X CGAA AGGCAAUG	486	CATTGCCTT T TGGGGGTG	1731
3329	GCACCCCC CUGAUGAG X CGAA AAGGCAAU	487	ATTGCCTT T GGGGGTGC	1732
3346	ACCCCAGG CUGAUGAG X CGAA AUCCCAGG	488	CCTGGGAT C CCTGGGGT	1733
3355	AGCUUGGAC CUGAUGAG X CGAA ACCCCAGG	489	CCTGGGGT A GTCCAGCT	1734
3358	AAGAGCUG CUGAUGAG X CGAA ACUACCCC	490	GGGGTAGT C CAGCTCTT	1735
3364	AUGAAUAA CUGAUGAG X CGAA AGCUGGAC	491	GTCCAGCT C TTATTAT	1736
3366	AAAUGAAU CUGAUGAG X CGAA AGAGCUGG	492	CCAGCTCTT ATTCAATT	1737

3367	GAAAUGAA CUGAUGAG X CGAA AAGAGCUG	493	CAGCTCTT A TTCATTTTC	1738
3369	GGGAAAUG CUGAUGAG X CGAA AUAGAGC	494	GCTCTTATT C ATTTCCC	1739
3370	UGGGAAAU CUGAUGAG X CGAA AAUAAGAG	495	CTCTTATT C ATTTCCC	1740
5	CGCUGGGA CUGAUGAG X CGAA AUGAAUAA	496	TTATTCAATT T TCCCAGCG	1741
3374	ACGCUGGG CUGAUGAG X CGAA AAUGAAUA	497	TATTCATT T CCCAGCGT	1742
3375	CACGCUGG CUGAUGAG X CGAA AAAUGAAU	498	ATTCATTT C CCAGCGTG	1743
3392	CUUCUUCC CUGAUGAG X CGAA ACCAGGGC	499	GCCCTGGTT GGAAGAAG	1744
10	3408 UACAACUU CUGAUGAG X CGAA ACAGCUGC	500	GCAGCTGT C AAGTTGTA	1745
3413	CUGUCUAC CUGAUGAG X CGAA ACUUGACA	501	TGTCAAGT T GTAGACAG	1746
3416	CAGCUGUC CUGAUGAG X CGAA ACAACUUG	502	CAAGTTGT A GACAGCTG	1747
3428	AUUGUAGG CUGAUGAG X CGAA ACACAGCU	503	AGCTGTGT T CCTACAAT	1748
15	3429 AAUUGUAG CUGAUGAG X CGAA AACACAGC	504	GCTGTGTT C CTACAATT	1749
3432	GCCAAUUG CUGAUGAG X CGAA AGGAACAC	505	GTGTTCTT A CAATTGGC	1750
3437	GCUGGGCC CUGAUGAG X CGAA AUUGUAGG	506	CCTACAATT GGCCCAGC	1751
3478	GUGACAGC CUGAUGAG X CGAA ACGGUCCC	507	GGGACCGTT GCTGTCAC	1752
3484	UGAGUAGU CUGAUGAG X CGAA ACAGCAAC	508	GTTGCTGT C ACTACTCA	1753
20	3488 AGCCUGAG CUGAUGAG X CGAA AGUGACAG	509	CTGTCACT A CTCAGGCT	1754
3491	GUCAGCCU CUGAUGAG X CGAA AGUAGUGA	510	TCACTACT C AGGCTGAC	1755
3511	CGUAAUCU CUGAUGAG X CGAA ACCAGGCC	511	GGCCTGGT C AGATTACG	1756
3516	GCAUACGU CUGAUGAG X CGAA AUCUGACC	512	GGTCAGATT ACGTATGC	1757
25	3517 GGCAUACG CUGAUGAG X CGAA AAUCUGAC	513	GTCAGATT A CGTATGCC	1758
3521	CAAGGGCA CUGAUGAG X CGAA ACGUAAUC	514	GATTACGT A TGCCCTTG	1759
3528	AAACCACCC CUGAUGAG X CGAA AGGGCAUA	515	TATGCCCT T GGTGGTTT	1760
3535	UAUCUCUA CUGAUGAG X CGAA ACCACCAA	516	TTGGTGGTT TAGAGATA	1761
30	3536 UUAUCUCU CUGAUGAG X CGAA AACCACCA	517	TGGTGGTT T AGAGATAA	1762
3537	AUUAAUCUC CUGAUGAG X CGAA AAACCACC	518	GGTGGTTT A GAGATAAT	1763
3543	UUUUGGAU CUGAUGAG X CGAA AUCUCUAA	519	TTAGAGAT A ATCCAAAAA	1764
3546	UGAUUUUG CUGAUGAG X CGAA AUUAUCUC	520	GAGATAAT C CAAATCA	1765
3553	CAAACCCU CUGAUGAG X CGAA AUUUUGGA	521	TCCAAAAT C AGGGTTTG	1766
3559	CCAAACCA CUGAUGAG X CGAA ACCCUGAU	522	ATCAGGGTT TGGTTTGG	1767
3560	CCCAAACC CUGAUGAG X CGAA AACCCUGA	523	TCAGGGTT GGTTGGG	1768

3564	CUUCCCCA CUGAUGAG X CGAA ACCAACCC	524	GGTTTGGTT T TGGGGAAG	1769
3565	UCUUCCCC CUGAUGAG X CGAA AACCAAAC	525	GTTTGGTT T GGGGAAGA	1770
3578	AGGGGGAG CUGAUGAG X CGAA AUUUUCUU	526	AAGAAAAT C CTCCCCCT	1771
3581	GGAAGGGG CUGAUGAG X CGAA AGGAUUUU	527	AAAATCCT C CCCCTTCC	1772
3587	GGGGGAGG CUGAUGAG X CGAA AGGGGGAG	528	CTCCCCCT T CCTCCCCC	1773
3588	CGGGGGAG CUGAUGAG X CGAA AAGGGGGA	529	TCCCCCTT C CTCCCCCG	1774
3591	GGGCGGGG CUGAUGAG X CGAA AGGAAGGG	530	CCCTTCCT C CCCCAGCC	1775
3603	CGGUAGGG CUGAUGAG X CGAA ACGGGGCG	531	CGCCCCGT T CCCTACCG	1776
3604	GCGGUAGG CUGAUGAG X CGAA AACGGGGC	532	GCCCCGTT C CCTACCGC	1777
3608	GGAGGCGG CUGAUGAG X CGAA AGGGAACG	533	CGTTCCCT A CCGCCCTCC	1778
3615	CAGGAGUG CUGAUGAG X CGAA AGGCGGUA	534	TACCGCCT C CACTCCTG	1779
3620	GCUGGCAG CUGAUGAG X CGAA AGUGGAGG	535	CCTCCACT C CTGCCAGC	1780
3630	AAGGAAAU CUGAUGAG X CGAA AGCUGGCA	536	TGCCAGCT C ATTTCCCT	1781
3633	UUGAAGGA CUGAUGAG X CGAA AUGAGCUG	537	CAGCTCAT T TCCTTCAA	1782
3634	AUUGAAGG CUGAUGAG X CGAA AAUGAGCU	538	AGCTCATT T CCTTCAAT	1783
3635	AAUUGAAG CUGAUGAG X CGAA AAAUGAGC	539	GCTCATTT C TTCAATT	1784
3638	GGAAAUUG CUGAUGAG X CGAA AGGAAAUG	540	CATTTCCCT T CAATTTCC	1785
3639	AGGAAAUU CUGAUGAG X CGAA AAGGAAAU	541	ATTTCCCT T CAATTTCC	1786
3643	UCAAAGGA CUGAUGAG X CGAA AUUGAAGG	542	CCTCAATT T TCCTTTGA	1787
3644	GUCAAAGG CUGAUGAG X CGAA AAUUGAAG	543	CTTCAATT T CCTTTGAC	1788
3645	GGUCAAAG CUGAUGAG X CGAA AAAUUGAA	544	TTCAATT T CCTTGTAC	1789
3648	AUAGGUCA CUGAUGAG X CGAA AGGAAAUU	545	AATTTCCCT T TGACCTAT	1790
3649	UAUAGGUC CUGAUGAG X CGAA AAGGAAAU	546	ATTTCCCT T GACCTATA	1791
3655	UUAGCCUA CUGAUGAG X CGAA AGGUAAA	547	TTTGACCT A TAGGCTAA	1792
3657	UUUUAGCC CUGAUGAG X CGAA AUAGGUCA	548	TGACCTAT A GGCTAAAA	1793
3662	UUCUUUUU CUGAUGAG X CGAA AGCCUAUA	549	TATAGGCT A AAAAAGAA	1794
3676	GCUGGAAU CUGAUGAG X CGAA AGCCUUUC	550	GAAAGGCT C ATTCCAGC	1795
3679	GUGGCUGG CUGAUGAG X CGAA AUGAGCCU	551	AGGCTCAT T CCAGCCAC	1796
3680	UGUGGCUG CUGAUGAG X CGAA AAUGAGCC	552	GGCTCAT T CAGCCACA	1797
3698	GCCCCAGG CUGAUGAG X CGAA AGGCUGCC	553	GGCAGCCT T CCCTGGGC	1798
3699	GGCCCAGG CUGAUGAG X CGAA AAGGCUGC	554	GCAGCCTT C CCTGGGCC	1799

5	3709	GAGAAGCA CUGAUGAG X CGAA AGGCCAG	555	CTGGGCCT T TGCTTCTC	1800
	3710	AGAGAAGC CUGAUGAG X CGAA AAGGCCA	556	TGGGCCTT T GCTCTCT	1801
	3714	UGCUAGAG CUGAUGAG X CGAA AGCAAAGG	557	CCTTGCT T CTCTAGCA	1802
	3715	GUGCUAGA CUGAUGAG X CGAA AAGCAAAG	558	CTTGCTT C TCTAGCAC	1803
	3717	UUGUGCUA CUGAUGAG X CGAA AGAAGCAA	559	TTGCTCT C TAGCACAA	1804
	3719	AAUUGUGC CUGAUGAG X CGAA AGAGAACG	560	GCTTCTCT A GCACAATT	1805
	3727	UAACCCAU CUGAUGAG X CGAA AUUGUGCU	561	AGCACAATT ATGGGTTA	1806
10	3728	GUAACCCA CUGAUGAG X CGAA AAUUGUGC	562	GCACAATT A TGGGTTAC	1807
	3734	AAGGAAGU CUGAUGAG X CGAA ACCCAUAA	563	TTATGGGT T ACTTCCTT	1808
	3735	AAAGGAAG CUGAUGAG X CGAA AACCCAUAA	564	TATGGGT A CTTCTTT	1809
	3738	GAAAAAGG CUGAUGAG X CGAA AGUAACCC	565	GGGTTACT T CCTTTTTC	1810
	3739	AGAAAAAG CUGAUGAG X CGAA AAGUAACC	566	GGTTACTT C CTTTTTCT	1811
15	3742	UUAAGAAA CUGAUGAG X CGAA AGGAAGUA	567	TACTTCCT T TTCTTAA	1812
	3743	GUUAAGAA CUGAUGAG X CGAA AAGGAAGU	568	ACTTCCTT T TTCTTAAC	1813
	3744	UGUUAAGA CUGAUGAG X CGAA AAAGGAAG	569	CTTCCTTT T TCTTAACA	1814
	3745	UUGUUAAG CUGAUGAG X CGAA AAAAGGAA	570	TTCCCTTT T CTTAACAA	1815
20	3746	UUUGUUAA CUGAUGAG X CGAA AAAAGGAA	571	TCCCTTTT C TTAACAAA	1816
	3748	UUUUUGUU CUGAUGAG X CGAA AGAAAAAG	572	CTTTTCTT AACAAAAAA	1817
	3749	UUUUUUGU CUGAUGAG X CGAA AAGAAAAAA	573	TTTTTCTT A ACAAAAAAA	1818
	3764	GGAAAUC CUGAUGAG X CGAA ACAUUCU	574	AAGAATGTT TGATTTCC	1819
25	3765	AGGAAAUC CUGAUGAG X CGAA AACAUUCU	575	AGAATGTT GATTTCT	1820
	3769	CCAGAGGA CUGAUGAG X CGAA AUCAAACA	576	TGTTTGATT TCCTCTGG	1821
	3770	CCCAGAGG CUGAUGAG X CGAA AAUCAAAC	577	GTITGATT CCTCTGGG	1822
	3771	ACCCAGAG CUGAUGAG X CGAA AAAUCAAA	578	TTTGATTT C CTCTGGGT	1823
	3774	GUCACCCA CUGAUGAG X CGAA AGGAAUC	579	GATTTCCT C TGGGTGAC	1824
30	3785	CAGACAAU CUGAUGAG X CGAA AGGUCACC	580	GGTGACCT T ATTGTCTG	1825
	3786	ACAGACAA CUGAUGAG X CGAA AAGGUCAC	581	GTGACCTT A TTGTCTGT	1826
	3788	UUACAGAC CUGAUGAG X CGAA AUAGGUC	582	GACCTTATT GTCTGTAA	1827
	3791	CAAUUACA CUGAUGAG X CGAA ACAAAUAG	583	CTTATTGT C TGTAATTG	1828
	3795	GUUUCAAU CUGAUGAG X CGAA ACAGACAA	584	TTGTCTGT A ATTGAAAC	1829
	3798	AGGGUUUC CUGAUGAG X CGAA AUUACAGA	585	TCTGTAATT GAAACCCT	1830

3807	CCUCUCAA CUGAUGAG X CGAA AGGUUUC	586	GAAACCCT A TTGAGAGG	1831
3809	CACCUCUC CUGAUGAG X CGAA AUAGGGUU	587	AACCCTATT GAGAGGTG	1832
3822	CUAACACA CUGAUGAG X CGAA ACAUCACC	588	GGTGATGT C TGTGTTAG	1833
3828	CAUUGGCU CUGAUGAG X CGAA ACACAGAC	589	GTCTGTGT T AGCCAATG	1834
3829	UCAUUGGC CUGAUGAG X CGAA AACACAGA	590	TCTGTGTT A GCCAATGA	1835
3845	CGAGCAGC CUGAUGAG X CGAA ACCUGGGU	591	ACCCAGGT A GCTGCTCG	1836
3852	AGAACCCC CUGAUGAG X CGAA AGCAGCUA	592	TAGCTGCT C GGGCTTCT	1837
3858	ACCAAGAG CUGAUGAG X CGAA AGCCCGAG	593	CTCGGGCT T CTCTTGGT	1838
3859	UACCAAGA CUGAUGAG X CGAA AAGCCCGA	594	TCGGGCTT C TCTTGGTA	1839
3861	CAUACCAA CUGAUGAG X CGAA AGAACCCC	595	GGGCTTCT C TTGGTATG	1840
3863	GACAUACC CUGAUGAG X CGAA AGAGAAC	596	GCTTCTCT T GGTATGTC	1841
3867	ACAAGACA CUGAUGAG X CGAA ACCAAGAG	597	CTCTTGGT A TGTCTTGT	1842
3871	CCAAACAA CUGAUGAG X CGAA ACAUACCA	598	TGGTATGT C TTGTTTGG	1843
3873	UUCCAAAC CUGAUGAG X CGAA AGACAUAC	599	GTATGTCT T GTTGGAA	1844
3876	CUUUUCCA CUGAUGAG X CGAA ACAAGACA	600	TGTCTTGT T TGGAAAAG	1845
3877	ACUUUUCC CUGAUGAG X CGAA AACAAAGAC	601	GTCTTGGT T GGAAAAGT	1846
3890	AUGAAUGA CUGAUGAG X CGAA AUCCACUU	602	AAGTGGATT TCATTTCAT	1847
3891	AAUGAAUG CUGAUGAG X CGAA AAUCCACU	603	AGTGGATT T CATTTCATT	1848
3892	AAAUGAAU CUGAUGAG X CGAA AAAUCCAC	604	GTGGATTTC ATTTCATT	1849
3895	CAGAAAUG CUGAUGAG X CGAA AUGAAAUC	605	GATTCATT CATTTCCTG	1850
3896	UCAGAAAU CUGAUGAG X CGAA AAUGAAA	606	ATTCATT C ATTTCCTGA	1851
3899	CAAUCAGA CUGAUGAG X CGAA AUGAAUGA	607	TCATTTCATT TCTGATTG	1852
3900	ACAAUCAG CUGAUGAG X CGAA AAUGAAUG	608	CATTTCATT T CTGATTGT	1853
3901	GACAAUCA CUGAUGAG X CGAA AAAUGAAU	609	ATTCATTTC TGATTGTC	1854
3906	AACUGGAC CUGAUGAG X CGAA AUCAGAAA	610	TTTCTGATT GTCCAGTT	1855
3909	CUUAACUG CUGAUGAG X CGAA ACAAUCAG	611	CTGATTGT C CAGTTAAG	1856
3914	GAUCACUU CUGAUGAG X CGAA ACUGGACA	612	TGTCCAGT T AAGTGATC	1857
3915	UGAUACACU CUGAUGAG X CGAA AACUGGAC	613	GTCCAGTT A AGTGATCA	1858
3922	CCUUUUGGU CUGAUGAG X CGAA AUCACUUA	614	TAAGTGATC ACCAAAGG	1859
3940	CCCUCCCCA CUGAUGAG X CGAA AUUCUCAG	615	CTGAGAATC TGGGAGGG	1860
3968	CACAUAAA CUGAUGAG X CGAA ACUUUUUU	616	AAAAAAAGTT TTATGTG	1861

3969	GCACAUAA CUGAUGAG X CGAA AACUUUUU	617	AAAAAGTT T TTATGTGC	1862
3970	UGCACAUAA CUGAUGAG X CGAA AACUUUUU	618	AAAAGTTT T TATGTGCA	1863
3971	GUGCACAU CUGAUGAG X CGAA AAAACUUU	619	AAAGTTTT T ATGTGCAC	1864
3972	AGUGCACA CUGAUGAG X CGAA AAAAACUU	620	AAGTTTTT A TGTGCACT	1865
3981	CCAAAUUU CUGAUGAG X CGAA AGUGCACA	621	TGTGCACT T AAATTGGA	1866
3982	CCCAAAUU CUGAUGAG X CGAA AAGUGCAC	622	GTGCACTT A AATTTGGG	1867
3986	UGUCCCCA CUGAUGAG X CGAA AUUUAAGU	623	ACTTAAATT T TGGGGACA	1868
3987	UUGUCCCC CUGAUGAG X CGAA AAUUAAG	624	CTTAAATT T GGGGACAA	1869
3997	AUACAUAA CUGAUGAG X CGAA AUUGUCCC	625	GGGACAATT T TTATGTAT	1870
3998	GAUACAUAA CUGAUGAG X CGAA AAUUGUCC	626	GGACAATT T TATGTATC	1871
3999	AGAUACAU CUGAUGAG X CGAA AAAUUGUC	627	GACAATT T TATGTATCT	1872
4000	CAGAUACA CUGAUGAG X CGAA AAAAUUGU	628	ACAATT T TGTATCTG	1873
4004	AACACAGA CUGAUGAG X CGAA ACAUAAAA	629	TTTTATGT A TCTGTGTT	1874
4006	UUAACACA CUGAUGAG X CGAA AUACAUAA	630	TTATGTATC TGTGTTAA	1875
4012	AUAUCCUU CUGAUGAG X CGAA ACACAGAU	631	ATCTGTGTT AAGGATAT	1876
4013	CAUAUCCU CUGAUGAG X CGAA AACACAGA	632	TCTGTGTT A AGGATATG	1877
4019	CUUAAGCA CUGAUGAG X CGAA AUCCUUA	633	TTAAGGAT A TGCTTAAG	1878
4024	AUGUUCUU CUGAUGAG X CGAA AGCAUAUC	634	GATATGCT T AAGAACAT	1879
4025	UAUGUUCU CUGAUGAG X CGAA AAGCAUAU	635	ATATGCTT A AGAACATA	1880
4033	AAAAGAAU CUGAUGAG X CGAA AUGUUCUU	636	AAGAACAT A ATTCTTT	1881
4036	ACACAAAG CUGAUGAG X CGAA AUUAUGUU	637	AACATAATT C TTTTGTT	1882
4037	CAACAAAA CUGAUGAG X CGAA AAUUAUGU	638	ACATAATT C TTTGTTG	1883
4039	AGCAACAA CUGAUGAG X CGAA AGAAUUAU	639	ATAATTCT T TTGTTGCT	1884
4040	CAGCAACA CUGAUGAG X CGAA AAGAAUUA	640	TAATTCTT T TGTTGCTG	1885
4041	ACAGCAAC CUGAUGAG X CGAA AAAGAAUU	641	AATTCTTT T GTTGTGT	1886
4044	CAAACAGC CUGAUGAG X CGAA ACAAAAGA	642	TCTTTGTT GCTGTTG	1887
4050	CUUAAACA CUGAUGAG X CGAA ACAGCAAC	643	GTGCTGTT T GTTAAAG	1888
4051	UCUAAAAC CUGAUGAG X CGAA AACAGCAA	644	TTGCTGTT T GTTAAAGA	1889
4054	GCUUCUUA CUGAUGAG X CGAA ACAAACAG	645	CTGTTGTT T TAAGAAC	1890
4055	UGCUUCUU CUGAUGAG X CGAA AACAAACA	646	TGTTGTT T AAGAAC	1891
4056	GUGCUUCU CUGAUGAG X CGAA AAACAAAC	647	GTGTTGTT A AGAAC	1892

5	4067	AACAAACU CUGAUGAG X CGAA AGGUGCUU	648	AAGCACCT T AGTTTGTT	1893
	4068	AAACAAAC CUGAUGAG X CGAA AAGGUGCU	649	AGCACCTT A GTTTGTTT	1894
	4071	CUUAAAACA CUGAUGAG X CGAA ACUAAGGU	650	ACCTTAGTT T GTTTAAG	1895
	4072	UCUUAAC CUGAUGAG X CGAA AACUAAGG	651	CCTTAGTT T GTTTAAGA	1896
	4075	GCUUCUUA CUGAUGAG X CGAA ACAAACUA	652	TAGTTTGT T TAAGAACG	1897
	4076	UGCUUCUU CUGAUGAG X CGAA AACAAACU	653	AGTTTGT T AAGAACGA	1898
	4077	GUGCUUCU CUGAUGAG X CGAA AAACAAAC	654	GTITGTIT T AGAACGAC	1899
10	4088	UACUUAU CUGAUGAG X CGAA AGGUGCU	655	AAGCACCT T ATATAGTA	1900
	4089	AUACUUA CUGAUGAG X CGAA AAGGUGCU	656	AGCACCTT A TATAGTAT	1901
	4091	UUAUACUA CUGAUGAG X CGAA AUAGGUG	657	CACCTTAT A TAGTATAA	1902
	4093	UAUUUAUAC CUGAUGAG X CGAA AUUAAGG	658	CCTTATAT A GTATAATA	1903
	4096	AUUAUUA CUGAUGAG X CGAA ACUUAUUA	659	TATATAGT A TAATATAT	1904
15	4098	AUUAUUAU CUGAUGAG X CGAA AUACUUA	660	TATAGTAT A ATATATAT	1905
	4101	AAAAUUA CUGAUGAG X CGAA AUUAUACU	661	AGTATAAT A TATATTTT	1906
	4103	AAAAAAUA CUGAUGAG X CGAA AUUAUUA	662	TATAATAT A TATTTTTT	1907
	4105	AAAAAAA CUGAUGAG X CGAA AUUAUUA	663	TAATATAT A TTTTTTGT	1908
20	4107	UUCAAAAA CUGAUGAG X CGAA AUUAUUAU	664	ATATATAT T TTTTGAA	1909
	4108	UUUCAAAA CUGAUGAG X CGAA AAUAUUA	665	TATATATT T TTTGAA	1910
	4109	AUUUCAA CUGAUGAG X CGAA AAAUUAU	666	ATATATT T TTTGAAAT	1911
	4110	AAUUUCAA CUGAUGAG X CGAA AAAAUUA	667	TATATTTT T TTGAAATT	1912
	4111	UAUUUUC CUGAUGAG X CGAA AAAAUUAU	668	ATATTTTT T TGAAATT	1913
25	4112	GUAAUUUC CUGAUGAG X CGAA AAAAUUA	669	TATTTTTT T GAAATTAC	1914
	4118	AGCAAUGU CUGAUGAG X CGAA AUUCAAA	670	TTGAAATT T ACATTGCT	1915
	4119	AAGCAAUG CUGAUGAG X CGAA AAUUUCAA	671	TTGAAATT A CATTGCTT	1916
	4123	AAACAAGC CUGAUGAG X CGAA AUGUAUU	672	AATTACATT GCTTGTTT	1917
30	4127	UGAUAAAAC CUGAUGAG X CGAA AGCAAUGU	673	ACATTGCT T GTTTATCA	1918
	4130	GUCUGAUA CUGAUGAG X CGAA ACAAGCAA	674	TTGCTTGTT T TATCAGAC	1919
	4131	UGUCUGAU CUGAUGAG X CGAA AACAGCA	675	TGCTTGTT T ATCAGACA	1920
	4132	UUGUCUGA CUGAUGAG X CGAA AAACAAGC	676	GCTTGTTT A TCAGACAA	1921
	4134	AAUUGUCU CUGAUGAG X CGAA AUAAACAA	677	TTGTTTAT C AGACAAATT	1922
	4142	CUACAUUC CUGAUGAG X CGAA AUUGUCUG	678	CAGACAATT GAATGTAG	1923

5	4149	AGAAUUC CUGAUGAG X CGAA ACAUUC	679	TTGAATGT A GTAATTCT	1924
	4152	AACAGAAU CUGAUGAG X CGAA ACUACAU	680	AATGTAGT A ATTCTGTT	1925
	4155	CAGAACAG CUGAUGAG X CGAA AUUACUAC	681	GTTAGTAATT CTGTTCTG	1926
	4156	CCAGAAC A CUGAUGAG X CGAA AAUUA	682	TAGTAATT C TGTCTGG	1927
	4160	AAAUCAG CUGAUGAG X CGAA ACAGAAU	683	AATTCTGTT CTGGATT	1928
	4161	AAAUCCA CUGAUGAG X CGAA AACAGAAU	684	ATTCTGTT C TGGATT	1929
	4167	UCAAAUUA CUGAUGAG X CGAA AUCCAGAA	685	TTCTGGATT TAATTG	1930
10	4168	GUCAAAUU CUGAUGAG X CGAA AAUCCAGA	686	TCTGGATT A ATTGAC	1931
	4169	AGUCAAAU CUGAUGAG X CGAA AAAUCCAG	687	CTGGATT A ATTGACT	1932
	4172	CCCAGUCA CUGAUGAG X CGAA AUUAAAUC	688	GATTAATT TGACTGGG	1933
	4173	ACCCAGUC CUGAUGAG X CGAA AAUAAA	689	ATTTAATT T GACTGGGT	1934
	4182	UGCAUGUU CUGAUGAG X CGAA ACCCAGUC	690	GACTGGTT AACATGCA	1935
15	4183	UUGCAUGU CUGAUGAG X CGAA AACCCAGU	691	ACTGGGTT A ACATGCAA	1936
	4207	AAACUAAA CUGAUGAG X CGAA AUUUUUC	692	GGAAAAAT A TTTAGTT	1937
	4209	AAAAACUA CUGAUGAG X CGAA AUUUUUU	693	AAAAATATT TAGTTTTT	1938
	4210	AAAAAACU CUGAUGAG X CGAA AAUAUUUU	694	AAAATATT T AGTTTTT	1939
20	4211	AAAAAAAC CUGAUGAG X CGAA AAAUAAA	695	AAATATT T GTTTTTT	1940
	4214	AAAAAAA CUGAUGAG X CGAA ACUAAA	696	TATTTAGTT TTTTTTT	1941
	4215	AAAAAAA CUGAUGAG X CGAA AACUAAA	697	ATTTAGTT TTTTTTT	1942
	4216	AAAAAAA CUGAUGAG X CGAA AAACAAA	698	TTTAGTTT TTTTTTT	1943
25	4217	AAAAAAA CUGAUGAG X CGAA AAAACUA	699	TTAGTTT TTTTTTT	1944
	4218	AAAAAAA CUGAUGAG X CGAA AAAACUA	700	TAGTTTT TTTTTTT	1945
	4219	AAAAAAA CUGAUGAG X CGAA AAAAACU	701	AGTTTTT TTTTTTT	1946
	4220	AAAAAAA CUGAUGAG X CGAA AAAAAC	702	GTTTTTT TTTTTTT	1947
	4221	AAAAAAA CUGAUGAG X CGAA AAAAAA	703	TTTTTTT TTTTTTT	1948
30	4222	CAAAAAAA CUGAUGAG X CGAA AAAAAA	704	TTTTTTT TTTTTTG	1949
	4223	ACAAAAAA CUGAUGAG X CGAA AAAAAA	705	TTTTTTT TTTTTGT	1950
	4224	UACAAAAA CUGAUGAG X CGAA AAAAAA	706	TTTTTTT TTTTGT	1951
	4225	AUACAAAA CUGAUGAG X CGAA AAAAAA	707	TTTTTTT TTTGTAT	1952
	4226	UAUACAAA CUGAUGAG X CGAA AAAAAA	708	TTTTTTT TTTGTATA	1953
	4227	GUAUACAA CUGAUGAG X CGAA AAAAAA	709	TTTTTTT T TTGTATAC	1954

4228	AGUAUACA CUGAUGAG X CGAA AAAAAAAA	710	TTTTTTTT T TGTATACT	1955
4229	AAGUAUAC CUGAUGAG X CGAA AAAAAAAA	711	TTTTTTTT T GTATACTT	1956
4232	GAAAAGUA CUGAUGAG X CGAA ACAAAAAA	712	TTTTTGT A TACTTTTC	1957
4234	UUGAAAAG CUGAUGAG X CGAA AUACAAAA	713	TTTTGTAT A CTTTCAA	1958
4237	AGCUUGAA CUGAUGAG X CGAA AGUAUACA	714	TGTATACT T TTCAAGCT	1959
4238	UAGCUUGA CUGAUGAG X CGAA AAGUAUAC	715	GTATACTTT TCAAGCTA	1960
4239	GUAGCUUG CUGAUGAG X CGAA AAAGUAUA	716	TATACTTT T CAAGCTAC	1961
4240	GGUAGCUU CUGAUGAG X CGAA AAAAGUAU	717	ATACTTTT C AAGCTACC	1962
4246	UGACALAGG CUGAUGAG X CGAA AGCUUGAA	718	TTCAAGCT A CCTTGTCA	1963
4250	UACAUAGAC CUGAUGAG X CGAA AGGUAGCU	719	AGCTACCTT GTCATGTA	1964
4253	GUAUACAU CUGAUGAG X CGAA ACAAGGUA	720	TACCTTGT C ATGTATAC	1965
4258	UGACUGUA CUGAUGAG X CGAA ACAUGACA	721	TGTATGT A TACAGTCA	1966
4260	AAUGACUG CUGAUGAG X CGAA AUACAUGA	722	TCATGTAT A CAGTCATT	1967
4265	GCAUAAA CUGAUGAG X CGAA ACUGUAUA	723	TATACAGT C ATTTATGC	1968
4268	UAGGCAUA CUGAUGAG X CGAA AUGACUGU	724	ACAGTCATT T ATGCCTA	1969
4269	UUAGGCAU CUGAUGAG X CGAA AAUGACUG	725	CAGTCATT T ATGCCTAA	1970
4270	UUUAGGCA CUGAUGAG X CGAA AAAUGACU	726	AGTCATT T TGCCTAAA	1971
4276	CCAGGCUU CUGAUGAG X CGAA AGGCAUAA	727	TTATGCCT A AAGCCTGG	1972
4289	AAAUGAAU CUGAUGAG X CGAA AUCACCAG	728	CTGGTGATT T ATTCAATT	1973
4290	UAAAUGAA CUGAUGAG X CGAA AAUCACCA	729	TGGTGATT A TTCAATT	1974
4292	UUUAAAUG CUGAUGAG X CGAA AUAUCAC	730	GTGATTAT T CATTTAAA	1975
4293	AUUUAAA CUGAUGAG X CGAA AAUAAUCA	731	TGATTATT C ATTTAAAT	1976
4296	UUCAUUUA CUGAUGAG X CGAA AUGAAUAA	732	TTATTCATT T TAAATGAA	1977
4297	CUUCAUU CUGAUGAG X CGAA AAUGAAUA	733	TATTCATT T AAATGAAG	1978
4298	UCUUCAUU CUGAUGAG X CGAA AAAUGAAU	734	ATTCATT T AATGAAGA	1979
4308	UGAAAAGU CUGAUGAG X CGAA AUCUUCAU	735	ATGAAGAT C ACATTCA	1980
4313	UGAUUAUGA CUGAUGAG X CGAA AUGUGAUC	736	GATCACATT TCATATCA	1981
4314	UUGAUUAUG CUGAUGAG X CGAA AAUGUGAU	737	ATCACATT T CATATCAA	1982
4315	GUUGAUAU CUGAUGAG X CGAA AAAUGUGA	738	TCACATTT C ATATCAAC	1983
4318	AAAGUUGA CUGAUGAG X CGAA AUGAAAUG	739	CATTTCAT A TCAACTTT	1984
4320	CAAAAGUU CUGAUGAG X CGAA AUAUGAAA	740	TTTCATAT C AACTTTG	1985

4325	GGAUACAA CUGAUGAG X CGAA AGUUGUA	741	TATCAACT T TTGTATCC	1986
4326	UGGAUACA CUGAUGAG X CGAA AAGUUGAU	742	ATCAACTT T TGTATCCA	1987
4327	GUGGAUAC CUGAUGAG X CGAA AAAGUUGA	743	TCAACTTT T GTATCCAC	1988
4330	ACUGUGGA CUGAUGAG X CGAA ACAAAAGU	744	ACTTTTGT A TCCACAGT	1989
4332	CUACUGUG CUGAUGAG X CGAA AUACAAAA	745	TTTTGTAT C CACAGTAG	1990
4339	AUUUUGUC CUGAUGAG X CGAA ACUGUGGA	746	TCCACAGT A GACAAAAT	1991
4348	AUUAGUGC CUGAUGAG X CGAA AUUUUGUC	747	GACAAAAT A GCACTAAT	1992
4354	AUCUGGAU CUGAUGAG X CGAA AGUGCUAU	748	ATAGCACT A ATCCAGAT	1993
4357	GGCAUCUG CUGAUGAG X CGAA AUUAGUGC	749	GCACTAAT C CAGATGCC	1994
4367	UCCAACAA CUGAUGAG X CGAA AGCCAUCU	750	AGATGCCT A TTGTTGGA	1995
4369	UAUCCAAC CUGAUGAG X CGAA AUAGGCAU	751	ATGCCTATT GTTGGATA	1996
4372	CAAAUACC CUGAUGAG X CGAA ACAAUAGG	752	CCTATTGTT GGATATTG	1997
4377	UCAUUCAA CUGAUGAG X CGAA AUCCAACA	753	TGTTGGAT A TTGAATGA	1998
4379	UGUCAUUC CUGAUGAG X CGAA AUAUCCAA	754	TTGGATATT GAATGACA	1999
4394	CUACAUAA CUGAUGAG X CGAA AUUGUCUG	755	CAGACAAT C TTATGTAG	2000
4396	UGCUACAU CUGAUGAG X CGAA AGAUUGUC	756	GACAATCTT ATGTAGCA	2001
4397	UUGCUACA CUGAUGAG X CGAA AAGAUUGU	757	ACAATCTT A TGTAGCAA	2002
4401	AUCUUUGC CUGAUGAG X CGAA ACAUAAGA	758	TCTTATGT A GCAAAGAT	2003
4410	UCAGGCCAU CUGAUGAG X CGAA AUCUUUGC	759	GCAAAGATT ATGCCTGA	2004
4411	UUCAGGCCA CUGAUGAG X CGAA AAUCUUUG	760	CAAAGATT A TGCCTGAA	2005
4429	CCCUGAAU CUGAUGAG X CGAA AUUUUCCU	761	AGGAAAATT ATTCAAGGG	2006
4430	GCCCUGAA CUGAUGAG X CGAA AAUUUUCC	762	GGAAAATT A TTCAGGGC	2007
4432	CUGCCCGUG CUGAUGAG X CGAA AUAAUUUU	763	AAAATTATT CAGGGCAG	2008
4433	GCUGCCCU CUGAUGAG X CGAA AAUAAUUU	764	AAATTATT C AGGGCAGC	2009
4443	AGCAAAAU CUGAUGAG X CGAA AGCUGCCC	765	GGGCAGCT A ATTITGCT	2010
4446	AAAAGCAA CUGAUGAG X CGAA AUUAGCUG	766	CAGCTAATT TTGCTTTT	2011
4447	AAAAAGCA CUGAUGAG X CGAA AAUUAGCU	767	AGCTAATT T TGCTTTTA	2012
4448	GUAAAAGC CUGAUGAG X CGAA AAAUUAGC	768	GCTAATT T GCTTTTAC	2013
4452	UUUGGUAA CUGAUGAG X CGAA AGCAAAAU	769	ATTITGCT T TTACCAAA	2014
4453	UUUUGGUAA CUGAUGAG X CGAA AAGCAAA	770	TTITGCTT T TACCAAAA	2015
4454	AUUUUGGU CUGAUGAG X CGAA AAAGCAAA	771	TTTGCTTT T ACCAAAAT	2016

5	4455	UAUUUUGG CUGAUGAG X CGAA AAAAGCAA	772	TTGCTTTT A CCAAAATA	2017
	4463	ACUACUGA CUGAUGAG X CGAA AUUUUGGU	773	ACCAAAAT A TCAGTAGT	2018
	4465	UUACUACU CUGAUGAG X CGAA AUAIUUUJUG	774	CAAAATAT C AGTAGTAA	2019
	4469	AAUAUUAC CUGAUGAG X CGAA ACUGAUAU	775	ATATCAGT A GTAATATT	2020
10	4472	AAAAAUAU CUGAUGAG X CGAA ACUACUGA	776	TCAGTAGT A ATATTTT	2021
	4475	UCCAAAAA CUGAUGAG X CGAA AUUACUAC	777	GTAGTAAT A TTTTGGAA	2022
	4477	UGUCCAAA CUGAUGAG X CGAA AUAUUACU	778	AGTAATATT T TTGGACACA	2023
	4478	CUGUCCAA CUGAUGAG X CGAA AAUAUUAC	779	GTAATATT T TTGGACAG	2024
15	4479	ACUGUCCA CUGAUGAG X CGAA AAAAUUUA	780	TAATATTT T TTGGACAGT	2025
	4480	UACUGUCC CUGAUGAG X CGAA AAAAUUUA	781	AATATTTT T GGACAGTA	2026
	4488	CCAUUAGC CUGAUGAG X CGAA ACUGUCCA	782	TGGACAGT A GCTAATGG	2027
	4492	UGACCCAU CUGAUGAG X CGAA AGCUACUG	783	CAGTAGCT A ATGGGTCA	2028
20	4499	AACCCACU CUGAUGAG X CGAA ACCCAUUA	784	TAATGGGT C AGTGGGTT	2029
	4507	UUAAAAAAG CUGAUGAG X CGAA ACCCACUG	785	CAGTGGGT T CTTTTTAA	2030
	4508	AUUAAAAA CUGAUGAG X CGAA AACCCACU	786	AGTGGGT C TTTTTAAT	2031
	4510	ACAUUAAA CUGAUGAG X CGAA AGAACCCA	787	TGGGTTCT T TTTAATGT	2032
25	4511	AACAUUAA CUGAUGAG X CGAA AAGAACCC	788	GGGTTCTT T TTAATGTT	2033
	4512	AAACAUUA CUGAUGAG X CGAA AAAGAACCC	789	GGTTCTTT T TAATGTTT	2034
	4513	UAAACAUU CUGAUGAG X CGAA AAAAGAAC	790	GTTCCTTT T AATGTTTA	2035
	4514	AUAAACAU CUGAUGAG X CGAA AAAAGAA	791	TTCTTTTT A ATGTTTAT	2036
30	4519	UAAGUAUA CUGAUGAG X CGAA ACAUAAA	792	TTTAATGT T TATACTTA	2037
	4520	CUAAGUAU CUGAUGAG X CGAA AACAUAAA	793	TTAATGTT T ATACCTAG	2038
	4521	UCUAAGUA CUGAUGAG X CGAA AAACAUUA	794	TAATGTT A TACTTACA	2039
	4523	AAUCUAAG CUGAUGAG X CGAA AUAAACAU	795	ATGTTTAT A CTTAGATT	2040
	4526	GAAAAUCU CUGAUGAG X CGAA AGUAAA	796	TTTATACCT T AGATTTTC	2041
	4527	AGAAAUC CUGAUGAG X CGAA AAGUAAA	797	TTATACCT A GATTTTCT	2042
	4531	UAAAAGAA CUGAUGAG X CGAA AUCUAAGU	798	ACTTAGAT T TTCTTTTA	2043
	4532	UUAAAAGA CUGAUGAG X CGAA AAUCUAAG	799	CTTAGATT T TCTTTTAA	2044
	4533	UUUAAAAG CUGAUGAG X CGAA AAAUCUAA	800	TTAGATTT T CTTTTAAA	2045
	4534	UUUUAAAA CUGAUGAG X CGAA AAAUCUA	801	TAGTTTCT C TTTTAAAA	2046
	4536	UUUUUUUA CUGAUGAG X CGAA AGAAAUC	802	GATTTTCT T TTAAAAAA	2047

5	4537	UUUUUUUA CUGAUGAG X CGAA AAGAAAAU	803	ATTTCTT T TAAAAAAA	2048
	4538	UUUUUUUU CUGAUGAG X CGAA AAAGAAAA	804	TTTCTTT T AAAAAAAT	2049
	4539	AAUUUUUU CUGAUGAG X CGAA AAAAGAAA	805	TTTCTTT A AAAAAATT	2050
	4547	UUUAUUUU CUGAUGAG X CGAA AUUUUUUU	806	AAAAAAAT T AAAATAAA	2051
10	4548	UUUUAUUU CUGAUGAG X CGAA AAUUUUUU	807	AAAAAAATT A AAATAAAA	2052
	4553	UUUUGUUU CUGAUGAG X CGAA AUUUUAAU	808	ATTAATAAAT A AAACAAAA	2053
	4567	GUCCUAGA CUGAUGAG X CGAA AUUUUUUU	809	AAAAAAATT T TCTAGGAC	2054
	4568	AGUCCUAG CUGAUGAG X CGAA AAUUUUUU	810	AAAAAATT T CTAGGACT	2055
15	4569	UAGUCCUA CUGAUGAG X CGAA AAAUUUUU	811	AAAAATTT C TAGGACTA	2056
	4571	UCUAGUCC CUGAUGAG X CGAA AGAAAUUU	812	AAATTTCT A GGACTAGA	2057
	4577	ACAUCGUC CUGAUGAG X CGAA AGUCCUAG	813	CTAGGACT A GACGATGT	2058
	4586	GCUGGUAU CUGAUGAG X CGAA ACAUCGUC	814	GACGATGT A ATACCAGC	2059
20	4589	UUAGCUGG CUGAUGAG X CGAA AUUACAUC	815	GATGTAAT A CCAGCTAA	2060
	4596	UUUGGCUU CUGAUGAG X CGAA AGCUGGUA	816	TACCAAGCT A AAGCCAAA	2061
	4609	CACUGUAU CUGAUGAG X CGAA AUUGUUUG	817	CAAACAATT ATACAGTG	2062
	4610	CCACUGUA CUGAUGAG X CGAA AAUUGUUU	818	AAACAATT A TACAGTGG	2063
25	4612	UUCCACUG CUGAUGAG X CGAA AUAAUUGU	819	ACAATTAT A CAGTGGAA	2064
	4624	UAAUGUAA CUGAUGAG X CGAA ACCUUCCA	820	TGGAAGGTT TTACATTA	2065
	4625	AUAAUGUA CUGAUGAG X CGAA AACCUUCC	821	GGAAGGTT T TACATTAT	2066
	4626	AAUAAUGU CUGAUGAG X CGAA AAACCUUC	822	GAAGGTTT T ACATTATT	2067
	4627	GAAUAAUG CUGAUGAG X CGAA AAAACCUU	823	AAGGTTTT A CATTATTC	2068
30	4631	GGAAUGAAU CUGAUGAG X CGAA AUGAAAA	824	TTTTACATT ATTCACTCC	2069
	4632	UGGAUGAA CUGAUGAG X CGAA AAUGUAAA	825	TTTACATT A TTCATCCA	2070
	4634	AUUGGAUG CUGAUGAG X CGAA AUAAUGUA	826	TACATTATT CATCCAAT	2071
	4635	CAUUGGAU CUGAUGAG X CGAA AAUAAUGU	827	ACATTATT C ATCCAATG	2072
	4638	ACACAUUG CUGAUGAG X CGAA AUGAAUAA	828	TTTACATT C CAATGTGT	2073
	4647	UGAAUAGA CUGAUGAG X CGAA ACACAUUG	829	CAATGTGT T TCTATTCA	2074
	4648	AUGAAUAG CUGAUGAG X CGAA AACACAUU	830	AATGTGT T CTATTCA	2075
	4649	CAUGAAUA CUGAUGAG X CGAA AAACACAU	831	ATGTGT T C TATTCA	2076
	4651	AAACAUAG CUGAUGAG X CGAA AGAAACAC	832	GTGTTTCT A TTCATGT	2077
	4653	UUAACAUG CUGAUGAG X CGAA AUAGAAAC	833	GTTCATATT CATGTAA	2078

4654	CUUAACA <del>U</del> CUGAUGAG X CGAA AAUAGAAA	834	TTTCTATT C ATGTTAAG	2079
4659	AGUAUCUU CUGAUGAG X CGAA ACAUGAAU	835	ATTCATGT T AAGATACT	2080
4660	UAGUAUCU CUGAUGAG X CGAA AACAUAGAA	836	TTCATGTT A AGATACTA	2081
4665	UGUAGUAG CUGAUGAG X CGAA AUCUUAAC	837	GTAAAGAT A CTACTACA	2082
4668	AAAUGUAG CUGAUGAG X CGAA AGUAUCUU	838	AAGATACT A CTACATT	2083
4671	UUCAAAUG CUGAUGAG X CGAA AGUAGUAU	839	ATACTACT A CATTGAA	2084
4675	CCACUUCA CUGAUGAG X CGAA AUGUAGUA	840	TACTACAT T TGAAGTGG	2085
4676	CCCACUUC CUGAUGAG X CGAA AAUGUAGU	841	ACTACATT T GAAGTGGG	2086
4695	AAUCAUCU CUGAUGAG X CGAA AUGUUCUC	842	GAGAACAT C AGATGATT	2087
4703	AACAUUUC CUGAUGAG X CGAA AUCAUCUG	843	CAGATGATT GAAATGTT	2088
4711	CCUGGGCG CUGAUGAG X CGAA ACAUUUCA	844	TGAAATGTT CGCCCAGG	2089
4712	CCCUGGGC CUGAUGAG X CGAA AACAUUUC	845	GAAATGTT C GCCCAGGG	2090
4723	UUGCUGGA CUGAUGAG X CGAA ACCCCUGG	846	CCAGGGGT C TCCAGCAA	2091
4725	AGUUGCUG CUGAUGAG X CGAA AGACCCU	847	AGGGGTCT C CAGCAACT	2092
4734	GAUUUCCA CUGAUGAG X CGAA AGUUGCUG	848	CAGCAACT T TGGAAATC	2093
4735	AGAUUUCC CUGAUGAG X CGAA AAGUUGCU	849	AGCAACTT T GGAAATCT	2094
4742	UACAAAGA CUGAUGAG X CGAA AUUCCAA	850	TTGGAAT C TCTTGTA	2095
4744	AAUACAAA CUGAUGAG X CGAA AGAUUUC	851	GGAAATCT C TTTGTATT	2096
4746	AAAAUACA CUGAUGAG X CGAA AGAGAUUU	852	AAATCTCT T TGTATTTT	2097
4747	AAAAAUAC CUGAUGAG X CGAA AAGAGAUU	853	AATCTCTT T GTATTTTT	2098
4750	AGUAAAAA CUGAUGAG X CGAA ACAAAGAG	854	CTCTTGT A TTTTACT	2099
4752	CAAGUAAA CUGAUGAG X CGAA AUACAAAG	855	C TTGTATT T TTACTTG	2100
4753	UCAAGUAA CUGAUGAG X CGAA AAUACAAA	856	TTTGTATT T TTACTTG	2101
4754	UUCAAGUA CUGAUGAG X CGAA AAAUACAA	857	TTGTATTT T TACTTGAA	2102
4755	CUUCAAGU CUGAUGAG X CGAA AAAAUACA	858	TGTATTTT T ACTTGAAAG	2103
4756	ACUUCAAG CUGAUGAG X CGAA AAAAUAC	859	GTATTTT A CTTGAAGT	2104
4759	GGCACUUC CUGAUGAG X CGAA AGUAAAAA	860	TTTTTACT T GAAGTGCC	2105
4771	CUGUCCAU CUGAUGAG X CGAA AGUGGCAC	861	GTGCCACT A ATGGACAG	2106
4785	CCAGAAAA CUGAUGAG X CGAA AUCUGCUG	862	CAGCAGAT A TTTCTGG	2107
4787	AGCCAGAA CUGAUGAG X CGAA AUAUCUGC	863	GCAGATATT TTCTGGCT	2108
4788	CAGCCAGA CUGAUGAG X CGAA AAUAUCUG	864	CAGATATT T TCTGGCTG	2109

4789	UCAGCCAG CUGAUGAG X CGAA AAAUAUCU	865	AGATATTT T CTGGCTGA	2110
4790	AUCAGCCA CUGAUGAG X CGAA AAAUAUC	866	GATATTTT C TGGCTGAT	2111
4801	CCAAUACC CUGAUGAG X CGAA ACAUCAGC	867	GCTGATGT T GGTATTGG	2112
4805	ACACCCAA CUGAUGAG X CGAA ACCAACAU	868	ATGTTGGT A TTGGGTGT	2113
4807	CUACACCC CUGAUGAG X CGAA AUACCAAC	869	GTGGTATT GGGTAG	2114
4814	CAUGUUCC CUGAUGAG X CGAA ACACCCAA	870	TTGGGTGT A GGAACATG	2115
4825	UUUUUUUA CUGAUGAG X CGAA AUCAUGUU	871	AACATGATT T AAAAAAAA	2116
4826	UUUUUUUU CUGAUGAG X CGAA AAUCAUGU	872	ACATGATT T AAAAAAAA	2117
4827	UUUUUUUU CUGAUGAG X CGAA AAAUCAUG	873	CATGATT T AAAAAAAA	2118
4839	AGAGGCAA CUGAUGAG X CGAA AGUUUUUU	874	AAAAAACT C TTGCCTCT	2119
4841	GCAGAGGC CUGAUGAG X CGAA AGAGUUUU	875	AAAACTCT T GCCTCTGC	2120
4846	GGAAAGCA CUGAUGAG X CGAA AGGCAAGA	876	TCTGCCT C TGCTTTCC	2121
4851	GUGGGGG CUGAUGAG X CGAA AGCAGAGG	877	CCTCTGCT T TCCCCCAC	2122
4852	AGUGGGGG CUGAUGAG X CGAA AAGCAGAG	878	CTCTGCTT T CCCCCACT	2123
4853	GAGUGGGG CUGAUGAG X CGAA AAAGCAGA	879	TCTGCTTC C CCCCCACTC	2124
4861	UUGCCUCA CUGAUGAG X CGAA AGUGGGGG	880	CCCCCACT C TGAGGCAA	2125
4872	UACAUUUU CUGAUGAG X CGAA ACUUGCCU	881	AGGCAAGT T AAAATGTA	2126
4873	UUACAUUU CUGAUGAG X CGAA AACUUGCC	882	GGCAAGTT A AAATGTA	2127
4880	ACAUCUUU CUGAUGAG X CGAA ACAUUUUA	883	TAAAATGT A AAAGATGT	2128
4892	CCCAGAUA CUGAUGAG X CGAA AUCACAU	884	GATGTGATT TATCTGGG	2129
4893	CCCCAGAU CUGAUGAG X CGAA AAUCACAU	885	ATGTGATT T ATCTGGGG	2130
4894	CCCCCAGA CUGAUGAG X CGAA AAAUCACA	886	TGTGATT T TCTGGGG	2131
4896	GCCCCCCA CUGAUGAG X CGAA AUAAAUC	887	TGATTTAT C TGGGGGGC	2132
4906	CCAUACCU CUGAUGAG X CGAA AGCCCCC	888	GGGGGGCT C AGGTATGG	2133
4911	CCCCACCA CUGAUGAG X CGAA ACCUGAGC	889	GCTCAGGT A TGGTGGGG	2134
4928	GAUCCUG CUGAUGAG X CGAA AUCCACUU	890	AAGTGGATT CAGGAATC	2135
4929	AGAUUCCU CUGAUGAG X CGAA AAUCCACU	891	AGTGGATT C AGGAATCT	2136
4936	AUUCCCCA CUGAUGAG X CGAA AUUCCUGA	892	TCAGGAAT C TGGGGAAAT	2137
4952	UCUUAAUA CUGAUGAG X CGAA AUUUGCCA	893	TGGCAAAT A TATTAAGA	2138
4954	CUUCUUA CUGAUGAG X CGAA AUAAUUGC	894	GCAAATAT A TTAAGAAG	2139
4956	CUCUUCUU CUGAUGAG X CGAA AUAAUUUU	895	AAATATATT AAGAAGAG	2140

4957	ACUCUUCU CUGAUGAG X CGAA AAUUAUU	896	AATATATT A AGAAGAGT	2141
4966	ACUUUCAA CUGAUGAG X CGAA ACUCUUCU	897	AGAAGAGT A TTGAAAGT	2142
4968	AUACUUUC CUGAUGAG X CGAA AUACUCUU	898	AAGAGTATT GAAAGTAT	2143
4975	CCUCCAAA CUGAUGAG X CGAA ACUUUCAA	899	TTGAAAGT A TTGGAGG	2144
4977	UUCCUCCA CUGAUGAG X CGAA AUACUUUC	900	GAAAGTATT TTGGAGGAA	2145
4978	UUUCCUCC CUGAUGAG X CGAA AAUACUUU	901	AAAGTATT GGAGGAAA	2146
4992	CCAGAAAU CUGAUGAG X CGAA ACCAUIUU	902	AAAATGGTT A ATTCTGG	2147
4993	CCCAGAAU CUGAUGAG X CGAA AACCAUUU	903	AAATGGTT A ATTCTGG	2148
4996	ACACCCAG CUGAUGAG X CGAA AUUAACCA	904	TGGTTAATT CTGGGTGT	2149
4997	CACACCCA CUGAUGAG X CGAA AAUUAACC	905	GCTTAATT CTGGGTGT	2150
5015	CUCUACUG CUGAUGAG X CGAA ACCUUGGU	906	ACCAAGGTT CAGTAGAG	2151
5016	ACUCUACU CUGAUGAG X CGAA AACCUUGG	907	CCAAGGTT C AGTAGAGT	2152
5020	GUGGACUC CUGAUGAG X CGAA ACUGAAC	908	GGTTCACT A GAGTCCAC	2153
5025	CAGAACUG CUGAUGAG X CGAA ACUCUACU	909	AGTAGAGT C CACTTCTG	2154
5030	CAGGGCAG CUGAUGAG X CGAA AGUGGACU	910	AGTCCACTT CTGCCCTG	2155
5031	CCAGGGCA CUGAUGAG X CGAA AAGUGGAC	911	GTCCACTT C TGCCCTGG	2156
5051	AGCUAGUU CUGAUGAG X CGAA AAUUGUGG	912	CCACAAAT C AACTAGCT	2157
5056	AAUGGAGC CUGAUGAG X CGAA AGUUGAUU	913	AATCAACT A GCTCCATT	2158
5060	UGUAAAUG CUGAUGAG X CGAA AGCUAGUU	914	AACTAGCT C CATTACAC	2159
5064	UGGCUGUA CUGAUGAG X CGAA AUGGAGCU	915	AGCTCCATT TACAGCCA	2160
5065	AUGGCUGU CUGAUGAG X CGAA AAUGGAGC	916	GCTCCATT T ACAGCCAT	2161
5066	AAUGGCUG CUGAUGAG X CGAA AAAUGGAG	917	CTCCATT T CAGCCATT	2162
5074	AUUUUAGA CUGAUGAG X CGAA AUGGCUGU	918	ACAGCCATT TCTAAAAT	2163
5075	CAUUUUAG CUGAUGAG X CGAA AAUGGCUG	919	CAGCCATT T CTAAAATG	2164
5076	CCAUUUUA CUGAUGAG X CGAA AAAUGGCU	920	AGCCATT T CTAAAATGG	2165
5078	UGCCAUUU CUGAUGAG X CGAA AGAAAUGG	921	CCATTCT A AAATGGCA	2166
5090	UAGAACUG CUGAUGAG X CGAA AGCUGCCA	922	TGGCAGCT T CAGITCTA	2167
5091	CUAGAACU CUGAUGAG X CGAA AAGCUGCC	923	GGCAGCTT C AGTTCTAG	2168
5095	UUCUCUAG CUGAUGAG X CGAA ACUGAAGC	924	GCTTCAGTT C TAGAGAGA	2169
5096	CUUCUCUA CUGAUGAG X CGAA AACUGAAG	925	CTTCAGTT C TAGAGAAG	2170
5098	UUCUUCUC CUGAUGAG X CGAA AGAACUGA	926	TCAGTTCT A GAGAAGAA	2171

5117	UUACUGCU CUGAUGAG X CGAA AUGUUGUU	927	AACAAACAT C ACCAGTAA	2172
5124	AUGGACUU CUGAUGAG X CGAA ACUGGUGA	928	TCAGCAGT A AAGTCCAT	2173
5129	AUUCCAUG CUGAUGAG X CGAA ACUUUACU	929	AGTAAAGT C CATGGAAT	2174
5138	CCACUAGC CUGAUGAG X CGAA AUUCCAUG	930	CATGGAAT A GCTAGTGG	2175
5142	CAGACCAC CUGAUGAG X CGAA AGCUAUUC	931	GAATAGCT A GTGGTCTG	2176
5148	GAAACACA CUGAUGAG X CGAA ACCACUAG	932	CTAGTGGT C TGTGTTTC	2177
5154	CGAAAAGA CUGAUGAG X CGAA ACACAGAC	933	GTCTGTGT T TCTTTTCG	2178
5155	GCGAAAAG CUGAUGAG X CGAA AACACAGA	934	TCTGTGTT T CTTTCGCG	2179
5156	GGCGAAAA CUGAUGAG X CGAA AACACACG	935	CTGTGTTT C TTTTCGCC	2180
5158	AUGGCGAA CUGAUGAG X CGAA AGAAACAC	936	GTGTTTCTT T TCGCCCAT	2181
5159	AAUGGCGA CUGAUGAG X CGAA AAGAAACA	937	TGTTTCTT T TCGCCATT	2182
5160	CAAUGGCC CUGAUGAG X CGAA AAAGAAC	938	GTTTCTTT T CGCCATTG	2183
5161	GCAAUGGC CUGAUGAG X CGAA AAAAGAAA	939	TTTCTTTT C GCCATTGC	2184
5167	AGCUAGGC CUGAUGAG X CGAA AUGGCGAA	940	TTCGCCATT GCCTAGCT	2185
5172	CGGCAAGC CUGAUGAG X CGAA AGGCAAUG	941	CATTGCCT A GCTTGCCG	2186
5176	AUUACGGC CUGAUGAG X CGAA AGCUAGGC	942	GCCTAGCT T GCCGTAAT	2187
5182	AGAAUCAU CUGAUGAG X CGAA ACGGCAAG	943	CTTGCCTG A ATGATTCT	2188
5188	CAUUUAUAG CUGAUGAG X CGAA AUCAUUAC	944	GTAATGATT CTATAATG	2189
5189	GCAUUUAU CUGAUGAG X CGAA AAUCAUUA	945	TAATGATT CTATAATGC	2190
5191	UGGCAUUUA CUGAUGAG X CGAA AGAAUCAU	946	ATGATTCT A TAATGCCA	2191
5193	GAUGGCAU CUGAUGAG X CGAA AUAGAAUC	947	GATTCTAT A ATGCCATC	2192
5201	UGCUGCAU CUGAUGAG X CGAA AUGGCAUU	948	AATGCCAT C ATGCAGCA	2193
5212	CCUCUCAU CUGAUGAG X CGAA AUUGCUGC	949	GCAGCAATT ATGAGAGG	2194
5213	GCCUCUCA CUGAUGAG X CGAA AAUUGCUG	950	CAGCAATT A TGAGAGGC	2195
5223	GGAUGACC CUGAUGAG X CGAA AGCCUCUC	951	GAGAGGCT A GGTCAATCC	2196
5227	CUUUUGAU CUGAUGAG X CGAA ACCUAGCC	952	GGCTAGGT C ATCCAAAG	2197
5230	UCUCUUUG CUGAUGAG X CGAA AUGACCUA	953	TAGGTCAT C CAAAGAGA	2198
5246	UACAUUGA CUGAUGAG X CGAA AGGGUCUU	954	AAGACCT A TCAATGTA	2199
5248	CCUACAUU CUGAUGAG X CGAA AUAGGGUC	955	GACCCTAT C AATGTAGG	2200
5254	UUGCAACC CUGAUGAG X CGAA ACAUUGAU	956	ATCAATGT A GGTTGCAA	2201
5258	GAUUUUGC CUGAUGAG X CGAA ACCUACAU	957	ATGTAGGT T GCAAAATC	2202

5266	AGGGGUUA CUGAUGAG X CGAA AUUUUGCA	958	TGCAAAAT C TAACCCCT	2203
5268	UUAGGGGU CUGAUGAG X CGAA AGAUUUUG	959	CAAAATCT A ACCCCTAA	2204
5275	CACUUCCU CUGAUGAG X CGAA AGGGGUUA	960	TAACCCCT A AGGAAGTG	2205
5288	AAAUCAAA CUGAUGAG X CGAA ACUGCACU	961	AGTGCAGT C TTTGATTT	2206
5290	UCAAAUCA CUGAUGAG X CGAA AGACUGCA	962	TGCAGTCT T TGATTTGA	2207
5291	AUCAAAUC CUGAUGAG X CGAA AAGACUGC	963	GCAGTCCT T GATTTGAT	2208
5295	GGAAAUCU CUGAUGAG X CGAA AUCAAAGA	964	TCTTGTAT T TGATTTCC	2209
5296	GGGAAACU CUGAUGAG X CGAA AAUCAAAG	965	CTTTGATT T GATTTCCC	2210
5300	ACUAGGG CUGAUGAG X CGAA AUCAAUC	966	GATTTGAT T TCCCTAGT	2211
5301	UACUAGGG CUGAUGAG X CGAA AAUCAAAU	967	ATTTGATT T CCCTAGTA	2212
5302	UUACUAGG CUGAUGAG X CGAA AAAUCAAA	968	TTTGATTT C CCTAGTAA	2213
5306	AAGGUUAC CUGAUGAG X CGAA AGGGAAAU	969	ATTTCCCT A GTAACCTT	2214
5309	UGCAAGGU CUGAUGAG X CGAA ACUAGGG	970	TCCCTAGT A ACCTTGCA	2215
5314	AUAUCUGC CUGAUGAG X CGAA AGGUUACU	971	AGTAACCT T GCAGATAT	2216
5321	GUUAAACA CUGAUGAG X CGAA AUCUGCAA	972	TTGCAGAT A TGTTAAC	2217
5325	CUUGGUUA CUGAUGAG X CGAA ACAUAUCU	973	AGATATGT T TAACCAAG	2218
5326	GCUUGGU CUGAUGAG X CGAA AACAUACU	974	GATATGTT AACCAAGC	2219
5327	GGCUUGGU CUGAUGAG X CGAA AAACAUAU	975	ATATGTTT A ACCAAGCC	2220
5338	GCAUGGGC CUGAUGAG X CGAA AUGGCUUG	976	CAAGCCAT A GCCCATGC	2221
5349	GCCCCUAA CUGAUGAG X CGAA AGGCAUGG	977	CCATGCCT T TTGAGGGC	2222
5350	AGCCCCU CUGAUGAG X CGAA AAGGCAUG	978	CATGCCTT T TGAGGGCT	2223
5351	CAGCCCUC CUGAUGAG X CGAA AAAGGCAU	979	ATGCCTTT T GAGGGCTG	2224
5367	AAGUCCCU CUGAUGAG X CGAA AUUUGUUC	980	GAACAAAT A AGGGACTT	2225
5375	UUAUCAGU CUGAUGAG X CGAA AGUCCUU	981	AAGGGACT T ACTGATAA	2226
5376	AUUAUCAG CUGAUGAG X CGAA AAGUCCCU	982	AGGGACTT A CTGATAAT	2227
5382	AAGUAAA CUGAUGAG X CGAA AUCAGUAA	983	TTACTGAT A ATTTACTT	2228
5385	CAAAAGUA CUGAUGAG X CGAA AUUAUCAG	984	CTGATAAT T TACTTTTG	2229
5386	UCAAAAGU CUGAUGAG X CGAA AAUUAUCA	985	TGATAATT T ACTTTTGA	2230
5387	AUCAAAAG CUGAUGAG X CGAA AAAUUAUC	986	GATAATT A CTTTGAT	2231
5390	GUGAUCAA CUGAUGAG X CGAA AGUAAA	987	AATTTACTT TTGATCAC	2232
5391	UGUGAUCA CUGAUGAG X CGAA AAGUAAA	988	ATTTACTT T TGATCAC	2233

5392	AUGUGAUC CUGAUGAG X CGAA AAAGUAAA	989	TTTACTTT T GATCACAT	2234
5396	CUUAAUGU CUGAUGAG X CGAA AUCAAAAG	990	CTTTGAT C ACATTAAG	2235
5401	AACACCUU CUGAUGAG X CGAA AUGUGAUC	991	GATCACAT T AAGGTGTT	2236
5402	GAACACCU CUGAUGAG X CGAA AAUGUGAU	992	ATCACATT A AGGTGTT	2237
5409	AAGGUGAG CUGAUGAG X CGAA ACACCUUA	993	TAAGGTGTT CTCACCTT	2238
5410	CAAGGUGA CUGAUGAG X CGAA AACACCUU	994	AAGGTGTT C TCACCTT	2239
5412	UUCAAGGU CUGAUGAG X CGAA AGAACACC	995	GGTGTCT C ACCTTGAA	2240
5417	AAGAUUUC CUGAUGAG X CGAA AGGUGAGA	996	TCTCACCT T GAAATCTT	2241
5423	GUGUAUAA CUGAUGAG X CGAA AUUUCAG	997	CTTGAAT C TTATACAC	2242
5425	CAGUGUAU CUGAUGAG X CGAA AGAUUUC	998	TGAAATCT T ATACACTG	2243
5426	UCAGUGUA CUGAUGAG X CGAA AAGAUUUC	999	GAAATCTT A TACACTGA	2244
5428	UUUCAGUG CUGAUGAG X CGAA AUAGAUU	1000	AATCTTAT A CACTGAAA	2245
5444	CCUAAAUC CUGAUGAG X CGAA AUGGCCAU	1001	ATGCCATT GATTTAGG	2246
5448	GUGGCCUA CUGAUGAG X CGAA AUCAAUGG	1002	CCATTGATT TAGGCCAC	2247
5449	AGUGGCCU CUGAUGAG X CGAA AAUCAAUG	1003	CATTGATT T AGGCCACT	2248
5450	CAGUGGCC CUGAUGAG X CGAA AAAUCAAU	1004	ATTGATT T GGCCACTG	2249
5462	AGUACUCU CUGAUGAG X CGAA AGCCAGUG	1005	CACTGGCT T AGAGTACT	2250
5463	GAGUACUC CUGAUGAG X CGAA AAGCCAGU	1006	ACTGGCTT A GAGTACTC	2251
5468	GGAAGGGAG CUGAUGAG X CGAA ACUCUAAG	1007	CTTAGAGT A CTCCCTCC	2252
5471	AGGGGAAG CUGAUGAG X CGAA AGUACUCU	1008	AGAGTACT C CTTCCCCT	2253
5474	UGCAGGGG CUGAUGAG X CGAA AGGAGUAC	1009	GTACTCCT T CCCCTGCA	2254
5475	AUGCAGGG CUGAUGAG X CGAA AAGGAGUA	1010	TACTCCTT C CCCTGCAT	2255
5493	GUAUUUGU CUGAUGAG X CGAA AUCAGUGU	1011	ACACTGATT ACAAAATAC	2256
5494	AGUAUUUG CUGAUGAG X CGAA AAUCAGUG	1012	CACTGATT A CAAATACT	2257
5500	UAGGAAAG CUGAUGAG X CGAA AUUUGUAA	1013	TTACAAAT A CTTCCTTA	2258
5503	GAUAGGGA CUGAUGAG X CGAA AGUAUUUG	1014	CAAATACT T TCCTATTTC	2259
5504	UGAAUAGG CUGAUGAG X CGAA AAGUAUUU	1015	AAATACCT T CCTATTCA	2260
5505	AUGAAUAG CUGAUGAG X CGAA AAAGUAAU	1016	AATACTTT C CTATTCA	2261
5508	AGUAUGAA CUGAUGAG X CGAA AGGAAAGU	1017	ACTTTCCCT A TTCATACT	2262
5510	AAAGUAUAG CUGAUGAG X CGAA AUAGGAAA	1018	TTTCCTATT T CATACTTT	2263
5511	GAAAGUAU CUGAUGAG X CGAA AAUAGGAA	1019	TTCCCTATT C ATACTTT	2264

5514	UUGGAAAG CUGAUGAG X CGAA AUGAAUAG	1020	CTATTCAT A CTTTCCAA	2265
5517	UAAUUGGA CUGAUGAG X CGAA AGUAUGAA	1021	TTCATACT T TCCAATTA	2266
5518	AUAAUUGG CUGAUGAG X CGAA AAGUAUGA	1022	TCATACTT T CCAATTAT	2267
5519	CAUAAUUG CUGAUGAG X CGAA AAAGUAUG	1023	CATACTTT C CAATTATG	2268
5524	CAUCUCAU CUGAUGAG X CGAA AUUGGAAA	1024	TTTCCAATT T ATGAGATG	2269
5525	CCAUCUCA CUGAUGAG X CGAA AAUUGGAA	1025	TTCCAATT A TGAGATGG	2270
5543	ACUCCCAG CUGAUGAG X CGAA ACCCACAG	1026	CTGTGGGT A CTGGGAGT	2271
5555	GUGUUAGU CUGAUGAG X CGAA AUCACUCC	1027	GGAGTGAT C ACTAACAC	2272
5559	UAUGGUGU CUGAUGAG X CGAA AGUGAUCA	1028	TGATCACT A ACACCATA	2273
5567	GACAUUAC CUGAUGAG X CGAA AUGGUGUU	1029	AACACCAT A GTAATGTC	2274
5570	UUAGACAU CUGAUGAG X CGAA ACUAUGGU	1030	ACCATAGT A ATGTCTAA	2275
5575	GAAUAAA CUGAUGAG X CGAA ACAUUACU	1031	AGTAATGT C TAATATTTC	2276
5577	GUGAAUAU CUGAUGAG X CGAA AGACAUUA	1032	TAATGTCT A ATATTCAC	2277
5580	CCUGUGAA CUGAUGAG X CGAA AUUAGACA	1033	TGTCTAAT A TTCACAGG	2278
5582	UGCCUGUG CUGAUGAG X CGAA AUUUAGA	1034	TCTAATATT CACAGGCA	2279
5583	CUGCCUGU CUGAUGAG X CGAA AAUAAUAG	1035	CTAATATT C ACAGGCAG	2280
5594	CCCAAGCA CUGAUGAG X CGAA AUCUGCCU	1036	AGGCAGAT C TGCTTGGG	2281
5599	GCUUCCCC CUGAUGAG X CGAA AGCAGAUC	1037	GATCTGCTT GGGGAAGC	2282
5609	CACAUAAAC CUGAUGAG X CGAA AGCUUCCC	1038	GGGAAGCT A GTTATGTG	2283
5612	UUUCACAU CUGAUGAG X CGAA ACUAGCUU	1039	AAAGCTAGT T ATGTGAAA	2284
5613	CUUUCACA CUGAUGAG X CGAA AACUAGCU	1040	AGCTAGTT A TGTGAAAG	2285
5628	UAUGACUU CUGAUGAG X CGAA AUUUGCCU	1041	AGGCAAAT A AAGTCATA	2286
5633	UACUGUAU CUGAUGAG X CGAA ACUUUAUU	1042	AATAAAGT C ATACAGTA	2287
5636	AGCUACUG CUGAUGAG X CGAA AUGACUUU	1043	AAAGTCAT A CAGTAGCT	2288
5641	UUUUGAGC CUGAUGAG X CGAA ACUGUAUG	1044	CATACAGT A GCTCAAAA	2289
5645	UGCCUUUU CUGAUGAG X CGAA AGCUACUG	1045	CAGTAGCT C AAAAGGCA	2290
5659	AAGAGAAU CUGAUGAG X CGAA AUGGUUGC	1046	GCAACCAT A ATTCTCTT	2291
5662	CCAAAGAG CUGAUGAG X CGAA AUUAUGGU	1047	ACCATAATT CTCTTTGG	2292
5663	ACCAAAGA CUGAUGAG X CGAA AAUUAUGG	1048	CCATAATT C TCTTTGGT	2293
5665	GCACCAAA CUGAUGAG X CGAA AGAAUUAU	1049	ATAATTCT C TTTGGTGC	2294
5667	UUGCACCA CUGAUGAG X CGAA AGAGAAUU	1050	AATTCTCTT TGGTGCAA	2295

5668	CUUGCACC CUGAUGAG X CGAA AAGAGAAU	1051	ATTCTCTT T GGTGCAAG	2296
5678	GCUCCCAA CUGAUGAG X CGAA ACUUGCAC	1052	GTGCAAGT C TTGGGAGC	2297
5680	ACGCUCCC CUGAUGAG X CGAA AGACUUGC	1053	GCAAGTCT T GGGAGCGT	2298
5692	GUAAUCUA CUGAUGAG X CGAA AUCACGCU	1054	AGCGTGAT C TAGATTAC	2299
5694	GUGUAUC CUGAUGAG X CGAA AGAUCACG	1055	CGTGATCT A GATTACAC	2300
5698	UGCAGUGU CUGAUGAG X CGAA AUCUAGAU	1056	ATCTAGATT ACAGTGCA	2301
5699	GUGCAGUG CUGAUGAG X CGAA AAUCUAGA	1057	TCTAGATT A CACTGCAC	2302
5711	AACUUGGG CUGAUGAG X CGAA AUGGUGCA	1058	TGCACCAT T CCCAAGTT	2303
5712	UAACUUGG CUGAUGAG X CGAA AAUGGUGC	1059	GCACCATTT C CCAAGTTA	2304
5719	AGGGGAUU CUGAUGAG X CGAA ACUUGGGA	1060	TCCCAAGTT AATCCCCCT	2305
5720	CAGGGGAU CUGAUGAG X CGAA AACUUGGG	1061	CCCAAGTT A ATCCCCGT	2306
5723	UUUCAGGG CUGAUGAG X CGAA AUUAACUU	1062	AAGTTAAT C CCCTGAAA	2307
5735	UUGAGAGU CUGAUGAG X CGAA AGUUUUCA	1063	TGAAAACCT T ACTCTCAA	2308
5736	GUUGAGAG CUGAUGAG X CGAA AAGUUUUC	1064	GAAAACCT A CTCTCAAC	2309
5739	CCAGUUGA CUGAUGAG X CGAA AGUAAGUU	1065	AACTTACT C TCAACTGG	2310
5741	CUCCAGUU CUGAUGAG X CGAA AGAGUAAG	1066	CTTACTCT C AACTGGAG	2311
5760	UGGGACCA CUGAUGAG X CGAA AGUUCAUU	1067	AATGAACCT T TGGTCCCA	2312
5761	UUGGGACC CUGAUGAG X CGAA AAGUUCAU	1068	ATGAACCTT GGTCCCAA	2313
5765	AUAUUJUGG CUGAUGAG X CGAA ACCAAAGU	1069	ACTTTGGT C CCAAATAT	2314
5772	AAGAUGGA CUGAUGAG X CGAA AUUUGGGA	1070	TCCCAAAT A TCCATCTT	2315
5774	AAAAGAUG CUGAUGAG X CGAA AUUUUUGG	1071	CCAAATAT C CATCTTTT	2316
5778	ACUGAAAA CUGAUGAG X CGAA AUGGAUAU	1072	ATATCCAT C TTTTCAGT	2317
5780	CUACUGAA CUGAUGAG X CGAA AGAUGGAU	1073	ATCCATCTT TTTCAGTAG	2318
5781	GCUACUGA CUGAUGAG X CGAA AAGAUGGA	1074	TCCATCTT T TCAGTAGC	2319
5782	CGCUACUG CUGAUGAG X CGAA AAAGAUGG	1075	CCATCTTT T CAGTAGCG	2320
5783	ACGCUACU CUGAUGAG X CGAA AAAAGAUG	1076	CATCTTT C AGTAGCGT	2321
5787	AUUAACGC CUGAUGAG X CGAA ACUGAAAA	1077	TTTTCACT A GCGTTAAT	2322
5792	GCAUAAUU CUGAUGAG X CGAA ACGCUACU	1078	AGTAGCGT T AATTATGC	2323
5793	AGCAUAAA CUGAUGAG X CGAA AACGCUAC	1079	GTAGCGTT A ATTATGCT	2324
5796	CAGAGCAU CUGAUGAG X CGAA AUUAACGC	1080	GCGTTAAT T ATGCTCTG	2325
5797	ACAGAGCA CUGAUGAG X CGAA AUUUAACG	1081	CGTTAATT A TGCTCTGT	2326

5802	UGGAAACA CUGAUGAG X CGAA AGCAUAAU	1082	ATTATGCT C TGTTTCCA	2327
5806	CAGUUGGA CUGAUGAG X CGAA ACAGAGCA	1083	TGCTCTGTT TCCAACTG	2328
5807	GCAGUUGG CUGAUGAG X CGAA AACAGAGC	1084	GCTCTGTT T CCAACTGC	2329
5808	UGCAGUUG CUGAUGAG X CGAA AACAGAG	1085	CTCTGTTT C CAACTGCA	2330
5818	GGAAAGGA CUGAUGAG X CGAA AUGCAGUU	1086	AACTGCATT T CCCTTTCC	2331
5819	UGGAAAGG CUGAUGAG X CGAA AAUGCAGU	1087	ACTGCATT T CCTTTCCA	2332
5820	UUGGAAAG CUGAUGAG X CGAA AAAUGCAG	1088	CTGCATT T CCTTTCCA	2333
5823	CAAUUGGA CUGAUGAG X CGAA AGGAAAUG	1089	CATTCCTT T TCCAATTG	2334
5824	UCAAUUGG CUGAUGAG X CGAA AAGGAAAU	1090	ATTCCTT T CCAATTGA	2335
5825	UUCAAUUG CUGAUGAG X CGAA AAAGGAAA	1091	TTTCCCTT C CAATTGAA	2336
5830	UUUAAUUC CUGAUGAG X CGAA AUUGGAAA	1092	TTTCCAATT GAATTAAA	2337
5835	CACACUUU CUGAUGAG X CGAA AUUCAAUU	1093	AATTGAATT AAAGTGTG	2338
5836	CCACACUU CUGAUGAG X CGAA AAUCAAU	1094	ATTGAATT A AAGTGTGG	2339
5848	CUAAAAAC CUGAUGAG X CGAA AGGCCACA	1095	TGTGGCCT C GTTTTAG	2340
5851	UGACUAAA CUGAUGAG X CGAA ACGAGGCC	1096	GGCCTCGTT TTTAGTCA	2341
5852	AUGACUAA CUGAUGAG X CGAA AACGAGGC	1097	GCCTCGTT T TTAGTCAT	2342
5853	AAUGACUA CUGAUGAG X CGAA AAACGAGG	1098	CCTCGTTT T TAGTCATT	2343
5854	AAAUGACU CUGAUGAG X CGAA AAAACGAG	1099	CTCGTTTT T AGTCATT	2344
5855	UAAAUGAC CUGAUGAG X CGAA AAAAACGA	1100	TCGTTTTT A GTCATTTA	2345
5858	UUUUAAA CUGAUGAG X CGAA ACUAAAAA	1101	TTTTTAGT C ATTTAAAA	2346
5861	CAAUUUUA CUGAUGAG X CGAA AUGACUAA	1102	TTAGTCATT T AAAATTG	2347
5862	ACAAUUUU CUGAUGAG X CGAA AAUGACUA	1103	TAGTCATT T AAAATTGT	2348
5863	ACAAUUUU CUGAUGAG X CGAA AAAUGACU	1104	AGTCATT T AAATTGTT	2349
5868	UAGAAAAC CUGAUGAG X CGAA AUUUUAAA	1105	TTTAAAATT T GTTTCTA	2350
5871	ACUUAGAA CUGAUGAG X CGAA ACAUUUU	1106	AAAATTGT T TTCTAAGT	2351
5872	UACUUAGA CUGAUGAG X CGAA AACAUUU	1107	AAATTGTT T TCTAAGTA	2352
5873	UUACUUAG CUGAUGAG X CGAA AAACAUU	1108	AATTGTTT T CTAAGTAA	2353
5874	AUUACUUA CUGAUGAG X CGAA AAAACAAU	1109	ATTGTTTT C TAAGTAAT	2354
5876	CAAUUACU CUGAUGAG X CGAA AGAAAACA	1110	TGTTTCT A AGTAATTG	2355
5880	GCAGCAAU CUGAUGAG X CGAA ACUUAGAA	1111	TTCTAAGT A ATTGCTGC	2356
5883	GAGGCAGC CUGAUGAG X CGAA AUUACUUA	1112	TAAGTAATT GCTGCCTC	2357

5891	CCAUAAUA CUGAUGAG X CGAA AGGCAGCA	1113	TGCTGCCT' C TATTATGG	2358
5893	UGCCAUAA CUGAUGAG X CGAA AGAGGCAG	1114	CTGCCTCT A TTATGGCA	2359
5895	AGUGCCAU CUGAUGAG X CGAA AUAGAGGC	1115	GCCTCTAT T ATGGCACT	2360
5896	AAGUGCCA CUGAUGAG X CGAA AAUAGAGG	1116	CCTCTATT A TGGCACTT	2361
5904	CAAAAUUG CUGAUGAG X CGAA AGUGCCAU	1117	ATGGCACT T CAATTTG	2362
5905	GCAAAAUU CUGAUGAG X CGAA AAGUGCCA	1118	TGGCACTT C AATTTTGC	2363
5909	CAGUGCAA CUGAUGAG X CGAA AUUGAAGU	1119	ACTTCAATT TTGCACTG	2364
5910	ACAGUGCA CUGAUGAG X CGAA AAUUGAAG	1120	CTTCAATT T TGCACGT	2365
5911	GACAGUGC CUGAUGAG X CGAA AAAUUGAA	1121	TTCAATT T GCACGTGTC	2366
5919	UCUCAAAA CUGAUGAG X CGAA ACAGUGCA	1122	TGCACTGT C TTTTGAGA	2367
5921	AAUCUCAA CUGAUGAG X CGAA AGACAGUG	1123	CACTGTCT T TTGAGATT	2368
5922	GAAUCUCA CUGAUGAG X CGAA AAGACAGU	1124	ACTGTCTT T TGAGATT	2369
5923	UGAAUCUC CUGAUGAG X CGAA AAAGACAG	1125	CTGTCTTT T GAGATTCA	2370
5929	UUUUCUUG CUGAUGAG X CGAA AUCUAAA	1126	TTTGGAGATT CAAGAAAA	2371
5930	UUUUUCUU CUGAUGAG X CGAA AAUCUAA	1127	TTGAGATT C AAGAAAAA	2372
5940	UGAAUAGA CUGAUGAG X CGAA AUUUUUCU	1128	AGAAAAATT T TCTATTCA	2373
5941	AUGAAUAG CUGAUGAG X CGAA AAUUUUUC	1129	GAAAAATT T CTATTCAT	2374
5942	AAUGAAUA CUGAUGAG X CGAA AAAUUUUU	1130	AAAAATT T CTATTCA	2375
5944	AAAUAUGAA CUGAUGAG X CGAA AGAAAUUU	1131	AAATTCT A TTCAATT	2376
5946	AAAAAAAUG CUGAUGAG X CGAA AUAGAAAU	1132	ATTTCTAT T CATTTTT	2377
5947	AAAAAAAU CUGAUGAG X CGAA AAUAGAAA	1133	TTTCTATT C ATTTTTT	2378
5950	UGCAAAAA CUGAUGAG X CGAA AUGAAUAG	1134	CTATTCAATT TTTTGCA	2379
5951	AUGCAAAA CUGAUGAG X CGAA AAUGAAUA	1135	TATTCAATT TTTGCAT	2380
5952	GAUGCAA CUGAUGAG X CGAA AAAUGAAU	1136	ATTCATTT T TTGCACTC	2381
5953	GGAUGCAA CUGAUGAG X CGAA AAAAUGAA	1137	TTCATTTT T TTGCACTCC	2382
5954	UGGAUGCA CUGAUGAG X CGAA AAAAUGA	1138	TCATTTTT T TGCACTCCA	2383
5955	UUGGAUGC CUGAUGAG X CGAA AAAAAUG	1139	CATTTTTT T GCATCCAA	2384
5960	CACAAUUG CUGAUGAG X CGAA AUGCAAAA	1140	TTTGCAT C CAATTGTG	2385
5965	UCAGGCAC CUGAUGAG X CGAA AUUGGAUG	1141	CATCCAATT GTGCCTGA	2386
5977	UAUUUUAA CUGAUGAG X CGAA AGUUCAGG	1142	CCTGAACATT TTAAAATA	2387
5978	AUAUUUUUA CUGAUGAG X CGAA AAGUUCAG	1143	CTGAACATT T AAAATAT	2388

2226	GCAGGA AGAA GAAU ACCAGAGAAACA X GUACAUUACCUGGUA	2551	ATTC TGTC TCCTGC	2664
2301	ACUAAG AGAA GAGC ACCAGAGAAACA X GUACAUUACCUGGUA	2552	GCTC AGTT CTTAGT	2665
2322	ACAGAA AGAA GAAG ACCAGAGAAACA X GUACAUUACCUGGUA	2553	CTTC TGTC TTCTGT	2666
2329	GUUCCC AGAA GAAG ACCAGAGAAACA X GUACAUUACCUGGUA	2554	CTTC TGTT GGGAAC	2667
2373	AAAGAG AGAA GUUA ACCAGAGAAACA X GUACAUUACCUGGUA	2555	TAAC AGCT CTCTTT	2668
2429	GAGUUC AGAA GUGA ACCAGAGAAACA X GUACAUUACCUGGUA	2556	TCAC AGCT GAACTC	2669
2439	CCCAUA AGAA GAGU ACCAGAGAAACA X GUACAUUACCUGGUA	2557	ACTC AGTC TATGGG	2670
2768	UAGGGG AGAA GCCU ACCAGAGAAACA X GUACAUUACCUGGUA	2558	AGGC AGAT CCCCTA	2671
2812	CUCUGA AGAA GCAG ACCAGAGAAACA X GUACAUUACCUGGUA	2559	CTGC AGAT TCAGAG	2672
2835	GCCAGA AGAA GAGC ACCAGAGAAACA X GUACAUUACCUGGUA	2560	GCTC TGCC TCTGGC	2673
2944	ACAAAA AGAA GGAA ACCAGAGAAACA X GUACAUUACCUGGUA	2561	TTCC TGAT TTTTGT	2674
3009	UCCUGA AGAA GACC ACCAGAGAAACA X GUACAUUACCUGGUA	2562	GGTC AGCT TCAGGA	2675
3021	CACUGG AGAA GGUC ACCAGAGAAACA X GUACAUUACCUGGUA	2563	GACC TGTT CCAGTG	2676
3083	ACAGUG AGAA GUUC ACCAGAGAAACA X GUACAUUACCUGGUA	2564	GAAC TGTT CACTGT	2677
3242	GCUCAG AGAA GUAU ACCAGAGAAACA X GUACAUUACCUGGUA	2565	ATAC AGTT CTGAGC	2678
3258	GAGCAA AGAA GGCU ACCAGAGAAACA X GUACAUUACCUGGUA	2566	AGCC AGAC TTGCTC	2679
3312	UGCGGG AGAA GCAA ACCAGAGAAACA X GUACAUUACCUGGUA	2567	TTGC AGAC CCCGCA	2680
3360	AAUAAG AGAA GGAC ACCAGAGAAACA X GUACAUUACCUGGUA	2568	GTCC AGCT CTTATT	2681
3402	CUUGAC AGAA GCUU ACCAGAGAAACA X GUACAUUACCUGGUA	2569	AAGC AGCT GTCAAG	2682
3420	GAACAC AGAA GUCU ACCAGAGAAACA X GUACAUUACCUGGUA	2570	AGAC AGCT GTGTT	2683
3475	GACAGC AGAA GUCC ACCAGAGAAACA X GUACAUUACCUGGUA	2571	GGAC CGTT GCTGTC	2684

3496	GCCCCA AGAA GCCU ACCAGAGAAACA X GUACAUUACCUGGU	2572	AGGC TGAC TGGGGC	2685
3512	UACGUA AGAA GACC ACCAGAGAAACA X GUACAUUACCUGGU	2573	GGTC AGAT TACGTA	2686
3595	GAACGG AGAA GGGG ACCAGAGAAACA X GUACAUUACCUGGU	2574	CCCC CGCC CCGTTC	2687
3600	GUAGGG AGAA GGGC ACCAGAGAAACA X GUACAUUACCUGGU	2575	CCCC CGTT CCCTAC	2688
3610	AGUGGA AGAA GUAG ACCAGAGAAACA X GUACAUUACCUGGU	2576	CTAC CGCC TCCACT	2689
3626	GAAAUG AGAA GGCA ACCAGAGAAACA X GUACAUUACCUGGU	2577	TGCC AGCT CATTTC	2690
3693	AGGGAA AGAA GCCC ACCAGAGAAACA X GUACAUUACCUGGU	2578	GGGC AGCC TTCCCT	2691
3848	AGCCCG AGAA GCUA ACCAGAGAAACA X GUACAUUACCUGGU	2579	TAGC TGCT CGGGCT	2692
3902	UGGACA AGAA GAAA ACCAGAGAAACA X GUACAUUACCUGGU	2580	TTTC TGAT TGTCCA	2693
4047	UAAACA AGAA GCAA ACCAGAGAAACA X GUACAUUACCUGGU	2581	TTGC TGTT TGTITA	2694
4157	AUCCAG AGAA GAAU ACCAGAGAAACA X GUACAUUACCUGGU	2582	ATTC TGTT CTGGAT	2695
4359	AUAGGC AGAA GGAU ACCAGAGAAACA X GUACAUUACCUGGU	2583	ATCC AGAT GCCTAT	2696
4696	UCAAUC AGAA GAUG ACCAGAGAAACA X GUACAUUACCUGGU	2584	CATC AGAT GATTGA	2697
4795	ACCAAC AGAA GCCA ACCAGAGAAACA X GUACAUUACCUGGU	2585	TGGC TGAT GTTGGT	2698
4847	GGGGAA AGAA GAGG ACCAGAGAAACA X GUACAUUACCUGGU	2586	CCTC TGCT TTCCCC	2699
5032	CUCCAG AGAA GAAG ACCAGAGAAACA X GUACAUUACCUGGU	2587	CTTC TGCC CTGGAG	2700
5086	AACUGA AGAA GCCA ACCAGAGAAACA X GUACAUUACCUGGU	2588	TGGC AGCT TCAGTT	2701
5092	CUCUAG AGAA GAAG ACCAGAGAAACA X GUACAUUACCUGGU	2589	CTTC AGTT CTAGAG	2702
5285	AUCAAA AGAA GCAC ACCAGAGAAACA X GUACAUUACCUGGU	2590	GTGC AGTC TTTGAT	2703
5489	UUUGUA AGAA GUGU ACCAGAGAAACA X GUACAUUACCUGGU	2591	ACACTGAT TACAAA	2704
5590	AAGCAG AGAA GCCU ACCAGAGAAACA X GUACAUUACCUGGU	2592	AGGC AGAT CTGCTT	2705

5595	UCCCCA AGAA GAUC ACCAGAGAAACA X GUACAUUACCUGGUA	2593	GATC TGCT TGGGGA	2706
5803	GUUGGA AGAA GAGC ACCAGAGAAACA X GUACAUUACCUGGUA	2594	GCTC TGTT TCCAAC	2707
5886	AAUAGA AGAA GCAA ACCAGAGAAACA X GUACAUUACCUGGUA	2595	TTGC TGCC TCTATT	2708
5916	UCAAAA AGAA GUGC ACCAGAGAAACA X GUACAUUACCUGGUA	2596	GCAC TGTC TTTGGA	2709
6087	AAAGGG AGAA GUGU ACCAGAGAAACA X GUACAUUACCUGGUA	2597	ACAC AGAC CCCTTT	2710
6154	AACAGA AGAA GGCA ACCAGAGAAACA X GUACAUUACCUGGUA	2598	TGCC AGTT TCTGTT	2711
6160	UGAGAG AGAA GAAA ACCAGAGAAACA X GUACAUUACCUGGUA	2599	TTTC TGTT CTCTCA	2712
6284	GUAUGC AGAA GCAA ACCAGAGAAACA X GUACAUUACCUGGUA	2600	TTGC CGAT GCATAC	2713
6300	AGUCAC AGAA GUAA ACCAGAGAAACA X GUACAUUACCUGGUA	2601	TTAC TGAT GTGACT	2714
6311	CGACAA AGAA GAGU ACCAGAGAAACA X GUACAUUACCUGGUA	2602	ACTC GGTT TTGTCG	2715
6322	AAGCAA AGAA GCGA ACCAGAGAAACA X GUACAUUACCUGGUA	2603	TCGC AGCT TTGCTT	2716

20

25

30

Table VI. Ribozymes for *in vitro* Cleavage

Seq. ID. No	Ribozyme Sequence	% CLEAVED ABOVE BACKGROUND
2727	[A <sub>s</sub> T <sub>s</sub> A <sub>s</sub> G <sub>s</sub> A <sub>s</sub> T <sub>s</sub> T <sub>s</sub> T <sub>s</sub> ] cUGAU	GaggccgaaaggccGaa Aggcacac B 3.2
2728	[G <sub>s</sub> C <sub>s</sub> G <sub>s</sub> G <sub>s</sub> A <sub>s</sub> C <sub>s</sub> C <sub>s</sub> ] cUGAU	GaggccgaaaggccGaa Agaaugaug B 11
2729	[T <sub>s</sub> T <sub>s</sub> T <sub>s</sub> C <sub>s</sub> C <sub>s</sub> G <sub>s</sub> A <sub>s</sub> ] cUGAU	GaggccgaaaggccGaa Agacaca B 1
2730	[A <sub>s</sub> T <sub>s</sub> T <sub>s</sub> C <sub>s</sub> C <sub>s</sub> TsGs] cUGAU	GaggccgaaaggccGaa Auuccuu B 80.8

Table VII. Antisense Nucleic Acid Molecules Targeting c-raf

Seq. I.D. No.	Sequence
2731	c <sub>s</sub> g <sub>s</sub> a <sub>s</sub> auugC <sub>s</sub> A <sub>s</sub> T <sub>s</sub> C <sub>s</sub> C <sub>s</sub> T <sub>s</sub> G <sub>s</sub> A <sub>s</sub> acag <sub>s</sub> a <sub>s</sub> A
2732	g <sub>s</sub> u <sub>s</sub> a <sub>s</sub> cctgA <sub>s</sub> T <sub>s</sub> T <sub>s</sub> C <sup>m</sup> <sub>s</sub> G <sub>s</sub> C <sub>s</sub> T <sub>s</sub> G <sub>s</sub> T <sub>s</sub> gacu <sub>s</sub> u <sub>s</sub> C <sub>s</sub> G
2733	g <sub>s</sub> C <sub>s</sub> a <sub>s</sub> ccagC <sub>s</sub> A <sub>s</sub> C <sub>s</sub> A <sub>s</sub> G <sub>s</sub> A <sub>s</sub> C <sub>s</sub> T <sub>s</sub> T <sub>s</sub> accu <sub>s</sub> g <sub>s</sub> a <sub>s</sub> T
2734	u <sub>s</sub> a <sub>s</sub> g <sub>s</sub> cagcC <sub>s</sub> C <sub>s</sub> T <sub>s</sub> G <sub>s</sub> A <sub>s</sub> G <sub>s</sub> C <sub>s</sub> C <sub>s</sub> T <sub>s</sub> uacc <sub>s</sub> u <sub>s</sub> g <sub>s</sub> G
2735	c <sub>s</sub> a <sub>s</sub> g <sub>s</sub> gauct <sub>s</sub> T <sub>s</sub> G <sub>s</sub> A <sub>s</sub> A <sub>s</sub> C <sub>s</sub> A <sub>s</sub> A <sub>s</sub> A <sub>s</sub> gccc <sub>s</sub> a <sub>s</sub> a <sub>s</sub> G
2736	u <sub>s</sub> g <sub>s</sub> C <sub>s</sub> cauct <sub>s</sub> T <sub>s</sub> T <sub>s</sub> A <sub>s</sub> C <sup>m</sup> <sub>s</sub> G <sub>s</sub> A <sub>s</sub> A <sub>s</sub> C <sub>s</sub> caac <sub>s</sub> C <sub>s</sub> C <sub>s</sub> A
2737	g <sub>s</sub> u <sub>s</sub> g <sub>s</sub> gtcaG <sub>s</sub> C <sup>m</sup> <sub>s</sub> G <sub>s</sub> T <sub>s</sub> G <sub>s</sub> A <sub>s</sub> G <sub>s</sub> cauu <sub>s</sub> g <sub>s</sub> a <sub>s</sub> T
2738	T <sub>s</sub> C <sub>s</sub> C <sub>s</sub> C <sub>s</sub> G <sub>s</sub> C <sub>s</sub> T <sub>s</sub> G <sub>s</sub> T <sub>s</sub> G <sub>s</sub> A <sub>s</sub> C <sub>s</sub> A <sub>s</sub> T <sub>s</sub> G <sub>s</sub> C <sub>s</sub> A <sub>s</sub> T <sub>s</sub>
2739	T <sub>s</sub> C <sub>s</sub> C <sub>s</sub> G <sub>s</sub> C <sub>s</sub> G <sub>s</sub> T <sub>s</sub> C <sub>s</sub> T <sub>s</sub> C <sub>s</sub> A <sub>s</sub> G <sub>s</sub> A <sub>s</sub> T <sub>s</sub> C <sub>s</sub> G <sub>s</sub> A <sub>s</sub> T <sub>s</sub>
2740	agucccgC <sub>s</sub> C <sub>s</sub> T <sub>s</sub> G <sub>s</sub> T <sub>s</sub> G <sub>s</sub> A <sub>s</sub> C <sub>s</sub> A <sub>s</sub> ugcauuc
2741	a <sub>s</sub> g <sub>s</sub> u <sub>s</sub> cccgC <sub>s</sub> C <sub>s</sub> T <sub>s</sub> G <sub>s</sub> T <sub>s</sub> G <sub>s</sub> A <sub>s</sub> C <sub>s</sub> A <sub>s</sub> ugcau <sub>s</sub> u <sub>s</sub> C <sub>s</sub>
2742	a <sub>s</sub> g <sub>s</sub> u <sub>s</sub> cccgC <sub>s</sub> C <sub>s</sub> T <sub>s</sub> G <sub>s</sub> T <sub>s</sub> G <sub>s</sub> A <sub>s</sub> C <sub>s</sub> A <sub>s</sub> ugcau <sub>s</sub> u <sub>s</sub> C <sub>s</sub>
2743	iB agucccgC <sub>s</sub> C <sub>s</sub> T <sub>s</sub> G <sub>s</sub> T <sub>s</sub> G <sub>s</sub> A <sub>s</sub> C <sub>s</sub> A <sub>s</sub> ugcauuc iB
2744	<b>agucccgC<sub>s</sub>C<sub>s</sub>T<sub>s</sub>G<sub>s</sub>T<sub>s</sub>G<sub>s</sub>A<sub>s</sub>C<sub>s</sub>A<sub>s</sub>ugcauuc</b> iB
2745	gauccgcC <sub>s</sub> G <sub>s</sub> T <sub>s</sub> C <sub>s</sub> T <sub>s</sub> C <sub>s</sub> A <sub>s</sub> G <sub>s</sub> A <sub>s</sub> ucgaucu
2746	g <sub>s</sub> a <sub>s</sub> u <sub>s</sub> ccgcC <sub>s</sub> G <sub>s</sub> T <sub>s</sub> C <sub>s</sub> T <sub>s</sub> C <sub>s</sub> A <sub>s</sub> G <sub>s</sub> A <sub>s</sub> ucga <sub>s</sub> u <sub>s</sub> C <sub>s</sub> u
2747	iB gauccgcC <sub>s</sub> G <sub>s</sub> T <sub>s</sub> C <sub>s</sub> T <sub>s</sub> C <sub>s</sub> A <sub>s</sub> G <sub>s</sub> A <sub>s</sub> ucgaucu iB
2748	<b>gauccgcC<sub>s</sub>G<sub>s</sub>T<sub>s</sub>C<sub>s</sub>T<sub>s</sub>C<sub>s</sub>A<sub>s</sub>G<sub>s</sub>A<sub>s</sub>ucgaucu</b> iB

lower case = 2'-O-methyl nucleotides; UPPER Case = DNA; s = phosphorothioate linkage;  
 m= 5 methyl C; iB=inverted abasic; **bold lower case**=2'-O-methylthiomethyl modified

Table VIII. Antisense Nucleic Acid Molecules Targeting Bcl-2 and K-ras

	Seq. I.D. No.	Target	Sequence
5	2749	Bcl-2	c <sub>s</sub> c <sub>s</sub> c <sub>s</sub> a <sub>s</sub> ccgA <sub>s</sub> A <sub>s</sub> C <sub>s</sub> T <sub>s</sub> C <sub>s</sub> A <sub>s</sub> A <sub>s</sub> A <sub>s</sub> G <sub>s</sub> aaggscscsa iB
	2750	Bcl-2	a <sub>s</sub> a <sub>s</sub> g <sub>s</sub> c <sub>s</sub> g <sub>a</sub> cC <sub>s</sub> T <sub>s</sub> A <sub>s</sub> A <sub>s</sub> G <sub>s</sub> C <sub>s</sub> A <sub>s</sub> A <sub>s</sub> ccccsa <sub>s</sub> g <sub>s</sub> c iB
	2751	Bcl-2	iB cccacccgA <sub>s</sub> A <sub>s</sub> C <sub>s</sub> T <sub>s</sub> C <sub>s</sub> A <sub>s</sub> A <sub>s</sub> A <sub>s</sub> G <sub>s</sub> aaggcca iB
	2752	Bcl-2	iBaaggcacC <sub>s</sub> T <sub>s</sub> A <sub>s</sub> A <sub>s</sub> A <sub>s</sub> G <sub>s</sub> C <sub>s</sub> A <sub>s</sub> A <sub>s</sub> cc cca gc iB
	2753	Bcl-2	c <sub>s</sub> c <sub>s</sub> c <sub>s</sub> a <sub>s</sub> ccgA <sub>s</sub> A <sub>s</sub> C <sub>s</sub> T <sub>s</sub> C <sub>s</sub> A <sub>s</sub> A <sub>s</sub> A <sub>s</sub> G <sub>s</sub> aa gg <sub>s</sub> c <sub>s</sub> c <sub>s</sub> a iB
	2754	k-ras (rat)	c <sub>s</sub> c <sub>s</sub> a <sub>s</sub> ccag C <sub>s</sub> T <sub>s</sub> C <sub>s</sub> C <sub>s</sub> A <sub>s</sub> A <sub>s</sub> C <sub>s</sub> tacc <sub>s</sub> a <sub>s</sub> c <sub>s</sub> A
10	2755	k-ras (rat)	t <sub>s</sub> g <sub>s</sub> g <sub>a</sub> caa a T <sub>s</sub> A <sub>s</sub> C <sub>s</sub> A <sub>s</sub> C <sub>s</sub> A <sub>s</sub> agaa <sub>s</sub> a <sub>s</sub> g <sub>s</sub> C
	2756	k-ras (rat)	c <sub>s</sub> c <sub>s</sub> a <sub>s</sub> taac T <sub>s</sub> C <sub>s</sub> C <sub>s</sub> T <sub>s</sub> T <sub>s</sub> G <sub>s</sub> C <sub>s</sub> taac <sub>s</sub> t <sub>s</sub> c <sub>s</sub> C
	2757	k-ras (rat)	c <sub>s</sub> a <sub>s</sub> c <sub>s</sub> cctg T <sub>s</sub> C <sub>s</sub> T <sub>s</sub> T <sub>s</sub> G <sub>s</sub> T <sub>s</sub> C <sub>s</sub> ttcg <sub>s</sub> c <sub>s</sub> t <sub>s</sub> G
	2758	Estrogen Receptor	c <sub>s</sub> u <sub>s</sub> g <sub>a</sub> ccagguT <sub>s</sub> G <sub>s</sub> G <sub>s</sub> T <sub>s</sub> C <sub>s</sub> A <sub>s</sub> G <sub>s</sub> uaagcc <sub>s</sub> c <sub>s</sub> u
15	2759	Estrogen Receptor	a <sub>s</sub> g <sub>s</sub> u <sub>s</sub> uuccaA <sub>s</sub> T <sub>s</sub> C <sub>s</sub> T <sub>s</sub> C <sub>s</sub> U <sub>s</sub> A <sub>s</sub> A <sub>s</sub> auug <sub>s</sub> g <sub>s</sub> c <sub>s</sub> a
	2760	Estrogen Receptor	T <sub>s</sub> G <sub>s</sub> C <sub>s</sub> <sup>M</sup> C <sub>s</sub> <sup>M</sup> A <sub>s</sub> G <sub>s</sub> G <sub>s</sub> T <sub>s</sub> T <sub>s</sub> G <sub>s</sub> G <sub>s</sub> T <sub>s</sub> C <sub>s</sub> <sup>M</sup> A <sub>s</sub> G <sub>s</sub> T <sub>s</sub> A <sub>s</sub> A <sub>s</sub> G <sub>s</sub> C <sub>s</sub> <sup>M</sup> C <sub>s</sub> <sup>M</sup> A <sub>s</sub>
	2761	Estrogen Receptor	T <sub>s</sub> C <sub>s</sub> <sup>M</sup> G <sub>s</sub> C <sub>s</sub> <sup>M</sup> A <sub>s</sub> T <sub>s</sub> G <sub>s</sub> T <sub>s</sub> G <sub>s</sub> C <sub>s</sub> <sup>M</sup> T <sub>s</sub> G <sub>s</sub> A <sub>s</sub> G <sub>s</sub> A <sub>s</sub> T <sub>s</sub> A <sub>s</sub> C <sub>s</sub> <sup>M</sup> G <sub>s</sub> C <sub>s</sub> <sup>M</sup> A <sub>s</sub>
	2762	Estrogen Receptor	ugc cag gT <sub>s</sub> T <sub>s</sub> G <sub>s</sub> G <sub>s</sub> T <sub>s</sub> C <sub>s</sub> A <sub>s</sub> G <sub>s</sub> T <sub>s</sub> aa gcc ca
20	2763	Estrogen Receptor	ucg cga uG <sub>s</sub> T <sub>s</sub> G <sub>s</sub> C <sub>s</sub> T <sub>s</sub> G <sub>s</sub> A <sub>s</sub> G <sub>s</sub> A <sub>s</sub> ua cgc ac
	2764	Estrogen Receptor	u <sub>s</sub> g <sub>s</sub> c <sub>s</sub> cag gT <sub>s</sub> T <sub>s</sub> G <sub>s</sub> G <sub>s</sub> T <sub>s</sub> C <sub>s</sub> A <sub>s</sub> G <sub>s</sub> T <sub>s</sub> aa gc <sub>s</sub> c <sub>s</sub> c <sub>s</sub> a
	2765	Estrogen Receptor	g <sub>s</sub> a <sub>s</sub> u <sub>s</sub> ccg cC <sub>s</sub> G <sub>s</sub> T <sub>s</sub> C <sub>s</sub> T <sub>s</sub> C <sub>s</sub> A <sub>s</sub> G <sub>s</sub> A <sub>s</sub> uc ga <sub>s</sub> u <sub>s</sub> c <sub>s</sub> u
	2766	Estrogen Receptor	iB ugc cag gT <sub>s</sub> T <sub>s</sub> G <sub>s</sub> G <sub>s</sub> T <sub>s</sub> C <sub>s</sub> A <sub>s</sub> G <sub>s</sub> T <sub>s</sub> aa gcc ca iB
25	2767	Estrogen Receptor	iB ucg cga uG <sub>s</sub> T <sub>s</sub> G <sub>s</sub> C <sub>s</sub> T <sub>s</sub> G <sub>s</sub> A <sub>s</sub> G <sub>s</sub> A <sub>s</sub> ua cgc ac iB
	2768	Estrogen Receptor	ugc cag gT <sub>s</sub> T <sub>s</sub> G <sub>s</sub> G <sub>s</sub> T <sub>s</sub> C <sub>s</sub> A <sub>s</sub> G <sub>s</sub> T <sub>s</sub> aa gcc ca iB
	2769	Estrogen Receptor	ucg cga u G <sub>s</sub> T <sub>s</sub> G <sub>s</sub> C <sub>s</sub> T <sub>s</sub> G <sub>s</sub> A <sub>s</sub> G <sub>s</sub> A <sub>s</sub> ua cgc ac iB
30			
35			

lower case = 2'-O-methyl nucleotides; UPPER Case = DNA; s = phosphorothioate linkage;  
**m** = 5 methyl C; iB = inverted abasic; **bold lower case** = 2'-O-methylthiomethyl modified

Claims

1. A nucleic acid molecule having the formula I:  
wherein each of X represents independently a nucleotide which may be same or different;



5 where m and o are integers independently greater than or equal to 5; (X)<sub>m</sub> and (X)<sub>o</sub> are oligonucleotides which are of sufficient length to stably interact independently with a target nucleic acid molecule; Y represents independently a deoxyribonucleotide which may be same or different; (Y)<sub>n</sub> is an oligonucleotide which is of sufficient length to stably interact independently with a target nucleic acid molecule; n is an integer greater than or  
10 equal to 4; \_\_\_\_\_ represents a chemical linkage; each (X)<sub>m</sub> and (X)<sub>o</sub> comprise independently at least one phosphodiester linkage and one phosphorothioate linkage; (Y)<sub>n</sub> comprises a phosphorothioate linkage or a phosphorodithioate linkage or a 5'-thiophosphate linkage or a mixture thereof; and each C and C' independently represents a cap structure which may independently be present or absent.

15 2. A nucleic acid molecule having the formula II:



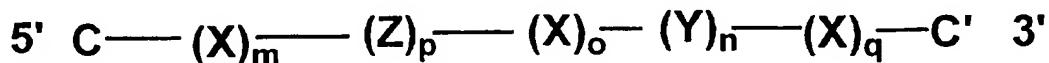
wherein X represents a nucleotide which may be same or different; where r is an integer greater than or equal to 4; (X)<sub>r</sub> is an oligonucleotide which is of sufficient length to stably interact independently with a target nucleic acid molecule; Y represents independently a deoxyribonucleotide which may be same or different; n is an integer greater than or equal to 4; (Y)<sub>n</sub> is an oligonucleotide which is of sufficient length to stably interact independently with a target nucleic acid molecule \_\_\_\_\_ represents a chemical linkage; (Y)<sub>n</sub> comprises a phosphorothioate linkage or a phosphorodithioate linkage or a 5'-thiophosphate linkage or a mixture thereof; each (X)<sub>r</sub> comprises independently at least one phosphodiester linkage and one phosphorothioate linkage; and each C and C' independently represents a cap structure which may independently be present or absent.

3. A nucleic acid molecule having the formula III:



wherein X represents a nucleotide which may be same or different; where r is an integer greater than or equal to 4;  $(\text{X})_r$  is an oligonucleotide which is of sufficient length to stably interact independently with a target nucleic acid molecule; Y represents independently a deoxyribonucleotide which may be same or different; n is an integer greater than or equal to 4;  $(\text{Y})_n$  is an oligonucleotide which is of sufficient length to stably interact independently with a target nucleic acid molecule \_\_\_\_\_ represents a chemical linkage;  $(\text{Y})_n$  comprises a phosphorothioate linkage or a phosphorodithioate linkage or a mixture of phosphorothioate and phosphorodithioate linkages; each  $(\text{X})_r$  comprises independently at least one phosphodiester linkage and one phosphorothioate linkage; and each C and C' independently represents a cap structure which may independently be present or absent.

4. An enzymatic nucleic acid molecule having endonuclease activity of the formula IV:



wherein each of X represents independently a nucleotide which may be same or different; where m, o and q are integers independently greater than or equal to 5;  $(\text{X})_m$  and  $(\text{X})_o$  are oligonucleotides which are of sufficient length to stably interact independently with a target nucleic acid molecule;  $(\text{X})_q$  is optionally able to interact with a target nucleic acid molecule;

Y represents independently a deoxyribonucleotide which may be same or different;  $(\text{Y})_n$  is an oligonucleotide which is of sufficient length to stably interact independently with a target nucleic acid molecule; n is an integer greater than or equal to 4; \_\_\_\_\_ represents a chemical linkage; each  $(\text{X})_m$ ,  $(\text{X})_o$  and  $(\text{X})_q$  comprise independently at least one phosphodiester linkage;  $(\text{Y})_n$  comprises a phosphorothioate linkage or a phosphorodithioate linkage or a 5'-thiophosphate linkage or a mixture thereof; Z represents an oligonucleotide able to facilitate the cleavage of the target nucleic acid molecule; p is of length greater than or equal to 4; and each C and C' independently represents a cap structure which may independently be present or absent.

5. An enzymatic nucleic acid molecule having endonuclease activity of the formula V:



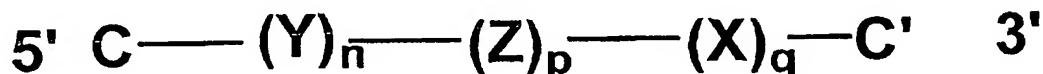
wherein each of X represents independently a nucleotide which may be same or different;

5 where m, o and q are integers independently greater than or equal to 5;  $(\text{X})_m$  and  $(\text{X})_o$  are oligonucleotides which are of sufficient length to stably interact independently with a target nucleic acid molecule;  $(\text{X})_q$  is optionally able to interact with the target nucleic acid molecule;  $(\text{Y})_n$  represents independently a deoxyribonucleotide which may be same or different;  $(\text{Y})_n$  is an oligonucleotide which is of sufficient length to stably interact

10 independently with a target nucleic acid molecule; n is an integer greater than or equal to 4; \_\_\_\_\_ represents a chemical linkage; each  $(\text{X})_m$ ,  $(\text{X})_o$  and  $(\text{X})_q$  comprise independently at least one phosphodiester linkage;  $(\text{Y})_n$  comprises a phosphorothioate linkage or a phosphorodithioate linkage or a 5'-thiophosphate linkage or a mixture thereof; Z represents an oligonucleotide able to facilitate the cleavage of the target nucleic acid

15 molecule; p is of length greater than or equal to 4; and each C and C' independently represents a cap structure which may independently be present or absent.

6. An enzymatic nucleic acid molecule having endonuclease activity of the formula VI:

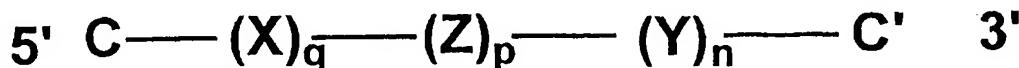


20 wherein X represents independently a nucleotide which may be same or different; where q is an integer independently greater than or equal to 1;  $(\text{X})_q$  is optionally able to interact with a target nucleic acid molecule; Y represents independently a deoxyribonucleotide which may be same or different;  $(\text{Y})_n$  is an oligonucleotide which is of sufficient length to stably interact independently with a target nucleic acid molecule; n is an integer greater

25 than or equal to 4; \_\_\_\_\_ represents a chemical linkage;  $(\text{Y})_n$  comprises a phosphorothioate linkage or a phosphorodithioate linkage or a 5'-thiophosphate linkage or a mixture thereof; Z represents an oligonucleotide able to facilitate the cleavage of the target nucleic acid

molecule; p is of length greater than or equal to 4; and each C and C' independently represents a cap structure which may independently be present or absent.

7. An enzymatic nucleic acid molecule having endonuclease activity of the formula VII:



5

wherein X represents independently a nucleotide which may be same or different; where q is an integer independently greater than or equal to 1; (X)<sub>q</sub> is optionally able to interact with a target nucleic acid molecule; Y represents independently a deoxyribonucleotide which may be same or different; (Y)<sub>n</sub> is an oligonucleotide which is of sufficient length to stably interact independently with the target nucleic acid molecule; n is an integer greater than or equal to 4; — represents a chemical linkage; (Y)<sub>n</sub> comprises a phosphorothioate linkage or a phosphorodithioate linkage or a 5'-S-phosphorothioate linkage or a 5'-S-phosphorodithioate linkage or a 3'-S-phosphorothioate linkage or a 3'-S-phosphorodithioate linkage or a mixture thereof; Z represents an oligonucleotide able to facilitate the cleavage of the target nucleic acid molecule; p is of length greater than or equal to 4; and each C and C' independently represents a cap structure which may independently be present or absent.

8. The nucleic acid molecule of any of claims 1-3, wherein each X, independently comprises a nucleotide modification selected from the group consisting of: 2'-O-methyl, 2'-O-allyl, 2'-O-methylthiomethyl, L-nucleotides; 2'-C-allyl; 1-5-Anhydrohexyitol; 2,6-diaminopurine; 2'-fluoro; 2'-deoxy-2'-amino; 2'-(N-alanyl) amino; 2'-(N-phenylalanyl) amino; 2'-deoxy-2'-(N-β-alanyl) amino; 2'-deoxy-2'-(lysyl) amino; 2'-O-amino; 2'-Deoxy-2'-(N-histidyl) amino; 6-methyl uridine; 5-methyl cytidine; 2'-(N-β-carboxamidine-β-alanyl) amino-2'-deoxy-nucleotide; 2'-O-methylthioallyl; 2'-O-methylthioethyl; 2'-O-methylthiomethyl; 2'-O-methyl-3'-thiophosphate and xylofuranosyl.

9. The enzymatic nucleic acid molecule of any of claims 4-7, wherein each X, Z or both X and Z, independently comprises a nucleotide modification selected from the group consisting of: 2'-O-methyl, 2'-O-allyl, 2'-O-methylthiomethyl, L-nucleotides; 2'-C-allyl; 1-5-Anhydrohexyitol; 2,6-diaminopurine; 2'-fluoro; 2'-deoxy-2'-amino; 2'-H; 2'-(N-

5 alanyl) amino; 2'-(*N*-phenylalanyl)amino; 2'-deoxy-2'-(*N*- $\beta$ -alanyl) amino; 2'-deoxy-2'-  
lysyl) amino; 2'-*O*-amino; 2'-Deoxy-2'-(*N*-histidyl) amino; 6-methyl uridine; 5-methyl  
cytidine; 2'-(*N*- $\beta$ -carboxamidine- $\beta$ -alanyl)amino-2'-deoxy-nucleotide; 2'-*O*-  
methylthioallyl; 2'-*O*-methylthioethyl; 2'-*O*-methylthiomethyl; 2'-*O*-methyl-3'-  
thiophosphate and xylofuranosyl.

10 10. The enzymatic nucleic acid molecule of any of claims 4-7, wherein, the Z  
in said nucleic acid molecule is the catalytic core.

11. 11. The enzymatic nucleic acid molecule of any of claims 4-7, wherein said  
enzymatic nucleic acid is in a hammerhead ribozyme configuration.

10 12. 12. The enzymatic nucleic acid molecule of any of claims 4-7, wherein said  
enzymatic nucleic acid is in a hairpin ribozyme configuration.

13. 13. The enzymatic nucleic acid of any of claims 4-7, wherein said enzymatic  
nucleic acid is in a hepatitis delta virus, group I intron, VS RNA, group II intron, or RNase  
P RNA configuration.

15 14. 14. A method of cleaving an RNA molecule comprising the step of, contacting  
the enzymatic nucleic acid molecule of any of claims 4-7, with the RNA molecule under  
conditions suitable for the cleavage of said RNA molecule by said enzymatic nucleic acid  
molecule.

15 15. 15. The method of claim 14, wherein said cleavage is carried out in the  
20 presence of a divalent cation.

16. 16. The method of claim 15, wherein said divalent cation is Mg<sup>2+</sup>.

17. 17. The nucleic acid molecules of any of claims 1-3, wherein said C', when  
present, is a cap selected from the group consisting of: inverted abasic residue; 4',5'-  
methylene nucleotide; 1-(beta-D-erythrofuranosyl) nucleotide; 4'-thio nucleotide,  
25 carbocyclic nucleotide; 1,5-anhydrohexitol nucleotide; L-nucleotides; alpha-nucleotides;  
modified base nucleotide; phosphorodithioate linkage; *threo*-pentofuranosyl nucleotide;  
acyclic 3',4'-seco nucleotide; acyclic 3,4-dihydroxybutyl nucleotide; acyclic 3,5-  
dihydroxypentyl nucleotide, 3'-3'-inverted nucleotide moiety; 3'-3'-inverted abasic moiety;

3'-2'-inverted nucleotide moiety; 3'-2'-inverted abasic moiety; 1,4-butanediol phosphate; 3'-phosphoramidate; hexylphosphate; aminohexyl phosphate; 3'-phosphate; 3'-phosphorothioate; phosphorodithioate; and methylphosphonate moiety.

18. The nucleic acid molecule of any of claim 1-3, wherein said nucleic acid  
5 molecule comprises a 3'-3' linked inverted ribose moiety at said 3' end.

19. The nucleic acid molecule of any of claims 1-3, wherein said nucleic acid  
molecule is an antisense nucleic acid molecule.

20. The nucleic acid molecule of claims 1-3 wherein said nucleic acid molecule  
is a 2-5A antisense chimera.

10 21. The nucleic acid molecule of any of claims 1-3 wherein said nucleic acid  
molecule is a triplex forming oligonucleotide.

22. A nucleic acid molecule comprising any of sequence defined as Seq. ID  
Nos. 1-30.

15 23. A mammalian cell including an enzymatic nucleic acid molecule of any of  
claims 4-7.

24. The mammalian cell of claim 23, wherein said mammalian cell is a human  
cell.

20 25. A method of modulating the expression of a gene in a cell comprising the  
step of administering to said cell a nucleic acid molecule of any of claims 1-3 under  
conditions suitable for the down regulation of said gene.

26. A method of modulating the expression of a gene in a cell comprising the  
step of administering to said cell an enzymatic nucleic acid molecule of any of claims 4-7  
under conditions suitable for the down regulation of said gene.

25 27. The enzymatic nucleic acid molecule of any of claims 4-7, wherein said C',  
when present, is a cap selected from the group consisting of: inverted abasic residue; 4',5'-  
methylene nucleotide; 1-(beta-D-erythofuranosyl) nucleotide; 4'-thio nucleotide,

carbocyclic nucleotide; 1,5-anhydrohexitol nucleotide; L-nucleotides; alpha-nucleotides; modified base nucleotide; phosphorodithioate linkage; *threo*-pentofuranosyl nucleotide; acyclic 3',4'-seco nucleotide; acyclic 3,4-dihydroxybutyl nucleotide; acyclic 3,5-dihydroxypentyl nucleotide, 3'-3'-inverted nucleotide moiety; 3'-3'-inverted abasic moiety; 5 3'-2'-inverted nucleotide moiety; 3'-2'-inverted abasic moiety; 1,4-butanediol phosphate; 3'-phosphoramidate; hexylphosphate; aminohexyl phosphate; 3'-phosphate; 3'-phosphorothioate; phosphorodithioate; and methylphosphonate moiety.

28. The enzymatic nucleic acid molecule of any of claims 4-7, wherein said C, when present, is a cap selected from the group consisting of: 4',5'-methylene nucleotide; 1-(beta-D-erythrophuranosyl) nucleotide; 4'-thio nucleotide, carbocyclic nucleotide; 5'-amino-alkyl phosphate; 1,3-diamino-2-propyl phosphate, 3-aminopropyl phosphate; 6-aminohexyl phosphate; 1,2-aminododecyl phosphate; hydroxypropyl phosphate; 1,5-anhydrohexitol nucleotide; L-nucleotide; alpha-nucleotide; modified base nucleotide; phosphorodithioate; *threo*-pentofuranosyl nucleotide; acyclic 3',4'-seco nucleotide; 3,4-dihydroxybutyl nucleotide; 3,5-dihydroxypentyl nucleotide, 5'-5'-inverted nucleotide moiety; 5'-5'-inverted abasic moiety; 5'-phosphoramidate; 5'-phosphorothioate; 1,4-butanediol phosphate; 5'-amino; bridging and/or non-bridging 5'-phosphoramidate, phosphorothioate and/or phosphorodithioate, bridging or non bridging methylphosphonate and 5'-mercapto moieties.

20 29. The nucleic acid molecules of any of claims 1-3, wherein said C, when present, is a cap selected from the group consisting of: 4',5'-methylene nucleotide; 1-(beta-D-erythrophuranosyl) nucleotide; 4'-thio nucleotide, carbocyclic nucleotide; 5'-amino-alkyl phosphate; 1,3-diamino-2-propyl phosphate, 3-aminopropyl phosphate; 6-aminohexyl phosphate; 1,2-aminododecyl phosphate; hydroxypropyl phosphate; 1,5-anhydrohexitol 25 nucleotide; L-nucleotide; alpha-nucleotide; modified base nucleotide; phosphorodithioate; *threo*-pentofuranosyl nucleotide; acyclic 3',4'-seco nucleotide; 3,4-dihydroxybutyl nucleotide; 3,5-dihydroxypentyl nucleotide, 5'-5'-inverted nucleotide moiety; 5'-5'-inverted abasic moiety; 5'-phosphoramidate; 5'-phosphorothioate; 1,4-butanediol phosphate; 5'-amino; bridging and/or non-bridging 5'-phosphoramidate, phosphorothioate and/or 30 phosphorodithioate, bridging or non bridging methylphosphonate and 5'-mercapto moieties.

30. A mammalian cell including a nucleic acid molecule of any of claims 1-3.

31. The mammalian cell of claim 30, wherein said mammalian cell is a human cell.

32. The enzymatic nucleic acid molecule of any of claims 4-7, wherein said enzymatic nucleic acid molecule comprises a 3'-3' linked inverted ribose moiety at said 3' end.

33. The nucleic acid molecule of any of claims 1-3, wherein said  $(X)_m$  and  $(X)_o$  are symmetric in length.

34. The nucleic acid molecule of any of claims 1-3, wherein said  $(X)_m$  and  $(X)_o$  are asymmetric in length.

10 35. The enzymatic nucleic acid molecule of any of claims 4-5, wherein said  $(X)_m$  and  $(X)_o$  are symmetric in length.

36. The enzymatic nucleic acid molecule of any of claims 4-5, wherein said  $(X)_m$  and  $(X)_o$  are asymmetric in length.

15 37. The enzymatic nucleic acid molecule of claim 4, wherein the sum of said o, n and q is equal to said m.

38. The enzymatic nucleic acid molecule of claim 5, wherein the sum of said o, n and m is equal to said q.

39. The enzymatic nucleic acid molecule of claim 4, wherein the sum of said o, n and q is greater than said m.

20 40. The enzymatic nucleic acid molecule of claim 5, wherein the sum of said o, n and m is greater than said q.

41. The enzymatic nucleic acid molecule of claim 4, wherein the sum of said o, n and q is less than said m.

25 42. The enzymatic nucleic acid molecule of claim 5, wherein the sum of said o, n and m is less than said q.

43. An enzymatic nucleic acid molecule with RNA cleaving activity, wherein said enzymatic nucleic acid molecule modulates the expression of an estrogen receptor gene.

44. The enzymatic nucleic acid molecule of claim 43, wherein said enzymatic nucleic acid molecule is in a hammerhead configuration.

45. The enzymatic nucleic acid molecule of claim 44, wherein said enzymatic nucleic acid molecule comprises a stem II region of length greater than or equal to 2 base pairs.

46. The enzymatic nucleic acid molecule of claim 43, wherein said enzymatic nucleic acid molecule is in a hairpin configuration.

47. The enzymatic nucleic acid molecule of claim 43, wherein said enzymatic nucleic acid is in a hepatitis d virus, group I intron, group II intron, VS nucleic acid or RNase P nucleic acid configuration.

48. The enzymatic nucleic acid of claim 46, wherein said enzymatic nucleic acid molecule comprises a stem II region of length between three and seven base-pairs.

49. The enzymatic nucleic acid molecule of claim 43, wherein said nucleic acid comprises between 12 and 100 bases complementary to said RNA.

50. The enzymatic nucleic acid molecule of claim 43, wherein said nucleic acid comprises between 14 and 24 bases complementary to said mRNA.

51. The enzymatic nucleic acid molecule of claim 44, wherein said enzymatic nucleic acid molecule consists essentially of any sequence defined as Seq ID Nos 1-1245.

52. A mammalian cell including an enzymatic nucleic acid molecule of any of claim 43.

53. The mammalian cell of claim 52, wherein said mammalian cell is a human cell.

54. An expression vector comprising nucleic acid sequence encoding at least one enzymatic nucleic acid molecule of claim 43, in a manner which allows expression of that enzymatic nucleic acid molecule.

55. A mammalian cell including an expression vector of claim 54.

5 56. The mammalian cell of claim 13, wherein said mammalian cell is a human cell.

57. A method for treatment of cancer comprising the step of administering to a patient the enzymatic nucleic acid molecule of claim 43.

10 58. A method for treatment of cancer comprising the step of administering to a patient the expression vector of claim 54.

59. A method for treatment of cancer comprising the steps of: a) isolating cells from a patient; b) administering to said cells the enzymatic nucleic acid molecule of claim 43; and c) introducing said cells back into said patient.

15 60. A pharmaceutical composition comprising the enzymatic nucleic acid molecule of claim 43.

61. A method of treatment of a patient having a condition associated with the level of estrogen receptor, wherein said patient is administered the enzymatic nucleic acid molecule of claim 43.

20 62. A method of treatment of a patient having a condition associated with the level of estrogen receptor, comprising contacting cells of said patient with the nucleic acid molecule of claim 43, and further comprising the use of one or more drug therapies.

25 63. The enzymatic nucleic acid molecule of claim 44, wherein said nucleic acid molecule comprises at least five ribose residues, and wherein said nucleic acid comprises phosphorothioate linkages at at least three of the 5' terminal nucleotides, and wherein said nucleic acid comprises a 2'-C-allyl modification at position No. 4 of said nucleic acid, and wherein said nucleic acid comprises at least ten 2'-O-methyl modifications, and wherein said nucleic acid comprises a 3'- end modification.

64. The enzymatic nucleic acid of claim 63, wherein said nucleic acid comprises a 3'-3' linked inverted ribose moiety at said 3' end.

5 65. The enzymatic nucleic acid molecule of claim 44, wherein said nucleic acid molecule comprises at least five ribose residues, and wherein said nucleic acid molecule comprises phosphorothioate linkages at at least three of the 5' terminal nucleotides, and wherein said nucleic acid comprises a 2'-amino modification at position No. 4 and/or at position No. 7 of said nucleic acid molecule, wherein said nucleic acid molecule comprises at least ten 2'-O-methyl modifications, and wherein said nucleic acid comprises a 3'- end modification.

10 66. The enzymatic nucleic acid molecule of claim 44, wherein said nucleic acid molecule comprises at least five ribose residues, and wherein said nucleic acid molecule comprises phosphorothioate linkages at at least three of the 5' terminal nucleotides, and wherein said nucleic acid molecule comprises an abasic substitution at position No. 4 and/or at position No. 7 of said nucleic acid molecule, wherein said nucleic acid comprises at least ten 2'-O-methyl modifications, and wherein said nucleic acid molecule comprises a 3'-end modification.

20 67. The enzymatic nucleic acid molecule of claim 44, wherein said nucleic acid molecule comprises of at least five ribose residues, and wherein said nucleic acid molecule comprises phosphorothioate linkages at at least three of the 5' terminal nucleotides, and wherein said nucleic acid molecule comprises a 6-methyl uridine substitution at position No. 4 and/or at position No. 7 of said nucleic acid molecule, wherein said nucleic acid molecule comprises at least ten 2'-O-methyl modifications, and wherein said nucleic acid molecule comprises a 3' end modification.

25 68. A method for modulating expression of estrogen receptor gene in a mammalian cell by administering to said cell the enzymatic nucleic acid molecule of claim 43.

69. A method of cleaving a separate RNA molecule comprising, contacting the enzymatic nucleic acid molecule of claim 43 with said separate RNA molecule under conditions suitable for the cleavage of said separate RNA molecule.

70. The method of claim 69, wherein said cleavage is carried out in the presence of a divalent cation.

71. The method of claim 70, wherein said divalent cation is  $Mg^{2+}$ .

72. The nucleic acid molecule of claim 43, wherein said nucleic acid is 5 chemically synthesized.

73. The expression vector of claim 54, wherein said vector comprises:

- a) a transcription initiation region;
- b) a transcription termination region;
- c) a gene encoding at least one said nucleic acid molecule; and

10 wherein said gene is operably linked to said initiation region and said termination region, in a manner which allows expression and/or delivery of said nucleic acid molecule.

74. The expression vector of claim 54, wherein said vector comprises:

- a) a transcription initiation region;
- b) a transcription termination region;
- c) an open reading frame;
- d) a gene encoding at least one said nucleic acid molecule, wherein said gene is operably linked to the 3'-end of said open reading frame; and

15 wherein said gene is operably linked to said initiation region, said open reading frame and said termination region, in a manner which allows expression and/or delivery of said nucleic acid molecule.

75. The expression vector of claim 54, wherein said vector comprises:

- a) a transcription initiation region;
- b) a transcription termination region;
- c) an intron;
- d) a gene encoding at least one said nucleic acid molecule; and

20 wherein said gene is operably linked to said initiation region, said intron and said termination region, in a manner which allows expression and/or delivery of said nucleic acid molecule.

30 76. The expression vector of claim 54, wherein said vector comprises:

- a) a transcription initiation region;
- b) a transcription termination region;
- c) an intron;
- d) an open reading frame;
- 5 e) a gene encoding at least one said nucleic acid molecule, wherein said gene is operably linked to the 3'-end of said open reading frame; and wherein said gene is operably linked to said initiation region, said intron, said open reading frame and said termination region, in a manner which allows expression and/or delivery of said nucleic acid molecule.

10 77. The enzymatic nucleic acid molecule of claim 44, wherein said enzymatic nucleic acid comprises sequences that are complementary to any of sequences defined as Seq ID Nos 1246-2490.

15 78. The enzymatic nucleic acid molecule of claim 46, wherein said enzymatic nucleic acid molecule consists essentially of any sequence defined as Seq ID Nos 2491-2603.

79. The enzymatic nucleic acid molecule of claim 46, wherein said enzymatic nucleic acid comprises sequences that are complementary to any of sequences defined as Seq ID Nos 2604-2716..

20 80. The enzymatic nucleic acid molecule of claim 43, wherein said enzymatic nucleic acid is a DNA enzyme.

81. The enzymatic nucleic acid molecule of claim 43, wherein said enzymatic nucleic acid comprises at least one 2'-sugar modification.

82. The enzymatic nucleic acid molecule of claim 43, wherein said enzymatic nucleic acid comprises at least one nucleic acid base modification.

25 83. The enzymatic nucleic acid molecule of claim 43, wherein said enzymatic nucleic acid comprises at least one phosphorothioate modification.

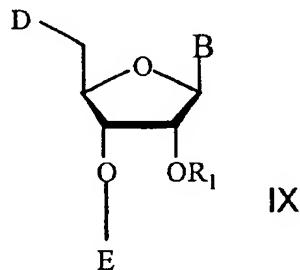
84. The enzymatic nucleic acid molecule of any of claims 4-7, wherein said enzymatic nucleic acid molecule comprises a 3'-3' linked inverted deoxyribose moiety at said 3' end.

5 85. The nucleic acid molecule of any of claims 1-3, wherein said enzymatic nucleic acid molecule comprises a 3'-3' linked inverted deoxyribose moiety at said 3' end.

86. The nucleic acid molecule of any of claims 1-3, wherein said enzymatic nucleic acid molecule comprises a 5'-5' linked inverted deoxyribose moiety at said 5' end.

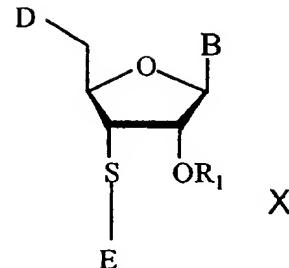
10 87. The enzymatic nucleic acid molecule of any of claims 4-7, wherein said enzymatic nucleic acid molecule comprises a 5'-5' linked inverted deoxyribose moiety at said 5' end.

88. The nucleic acid molecule of any of claims 1-3, wherein each X, independently comprises a nucleotide modification having formula IX:



15 Wherein, each B is independently a modified or an unmodified nucleic acid base; R1 is independently an alkyl, an alkylthioalkyl, a fluoroalkyl or an alkylthiofluoroalkyl; E is independently a phosphorus-containing group; and D is independently an O, blocking group or a phosphorus-containing group.

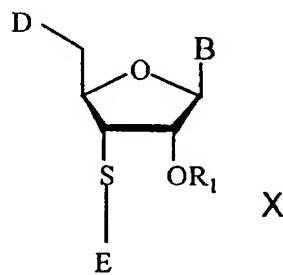
89. The nucleic acid molecule of any of claims 1-3, wherein each X, independently comprises a nucleotide modification having formula X:



Wherein, each B is independently a modified or an unmodified nucleic acid base; R<sub>1</sub> is independently an alkyl, an alkylthioalkyl, a fluoroalkyl or an alkylthiofluoroalkyl; E is independently a phosphorus-containing group; and D is independently an O, blocking group or a phosphorus-containing group.

5

90. The nucleic acid molecule of any of claims 4-7, wherein each X, Z or both X and Z, independently comprise a nucleotide modification having formula X:

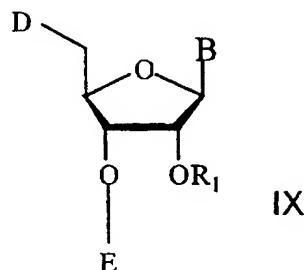


10

Wherein, each B is independently a modified or an unmodified nucleic acid base; R<sub>1</sub> is independently an alkyl, an alkylthioalkyl, a fluoroalkyl or an alkylthiofluoroalkyl; E is independently a phosphorus-containing group; and D is independently an O, blocking group or a phosphorus-containing group.

15

91. The nucleic acid molecule of any of claims 4-7, wherein each X, Z or both X and Z, independently comprises a nucleotide modification having formula IX:



Wherein, each B is independently a modified or an unmodified nucleic acid base; R1 is independently an alkyl, an alkylthioalkyl, a fluoroalkyl or an alkylthiofluoroalkyl; E is independently a phosphorus-containing group; and D is independently an O, blocking group or a phosphorus-containing group.

5           92. The nucleic acid molecule of any of claims 1-3 and 4-7, wherein  $(Y)_n$  comprises a phosphorothioate linkage.

10          93. The nucleic acid molecule of any of claims 1-3 and 4-7, wherein  $(Y)_n$  comprises a phosphorodithioate linkage.

15          94. The nucleic acid molecule of any of claims 1-3 and 4-7, wherein  $(Y)_n$  comprises a 5'-S-phosphorothioate linkage.

20          95. The nucleic acid molecule of any of claims 1-3 and 4-7, wherein  $(Y)_n$  consists of phosphorothioate linkage at every position.

15          96. The nucleic acid molecule of any of claims 1-3 and 4-7, wherein  $(Y)_n$  consists of phosphorodithioate linkage at every position.

25          97. The nucleic acid molecule of any of claims 1-3 and 4-7, wherein  $(Y)_n$  consists of 5'-S-phosphorothioate linkage at every position.

30          98. The nucleic acid molecule of any of claims 1-3 and 4-7, wherein  $(Y)_n$  comprises a combination of phosphorothioate and phosphorodithioate linkages.

35          99. The nucleic acid molecule of any of claims 1-3 and 4-7, wherein  $(Y)_n$  comprises a combination of phosphorothioate and 5'-S-phosphorothioate linkages.

100. The nucleic acid molecule of any of claims 1-3 and 4-7, wherein (Y)<sub>n</sub> comprises a combination of phosphorodithioate and 5'-S-phosphorothioate linkages.

101. The nucleic acid molecule of any of claims 1-3 and 4-7, wherein (Y)<sub>n</sub> comprises a combination of phosphorothioate, phosphorodithioate and 5'-S-phosphorothioate linkages.

102. The nucleic acid molecule of any of claims 1-3 and 4-7, wherein (Y)<sub>n</sub> comprises a combination of phosphorothioate and 5'-S-phosphorodithioate linkages.

103. The nucleic acid molecule of any of claims 1-3 and 4-7, wherein (Y)<sub>n</sub> comprises a combination of phosphorodithioate and 5'-S-phosphorodithioate linkages.

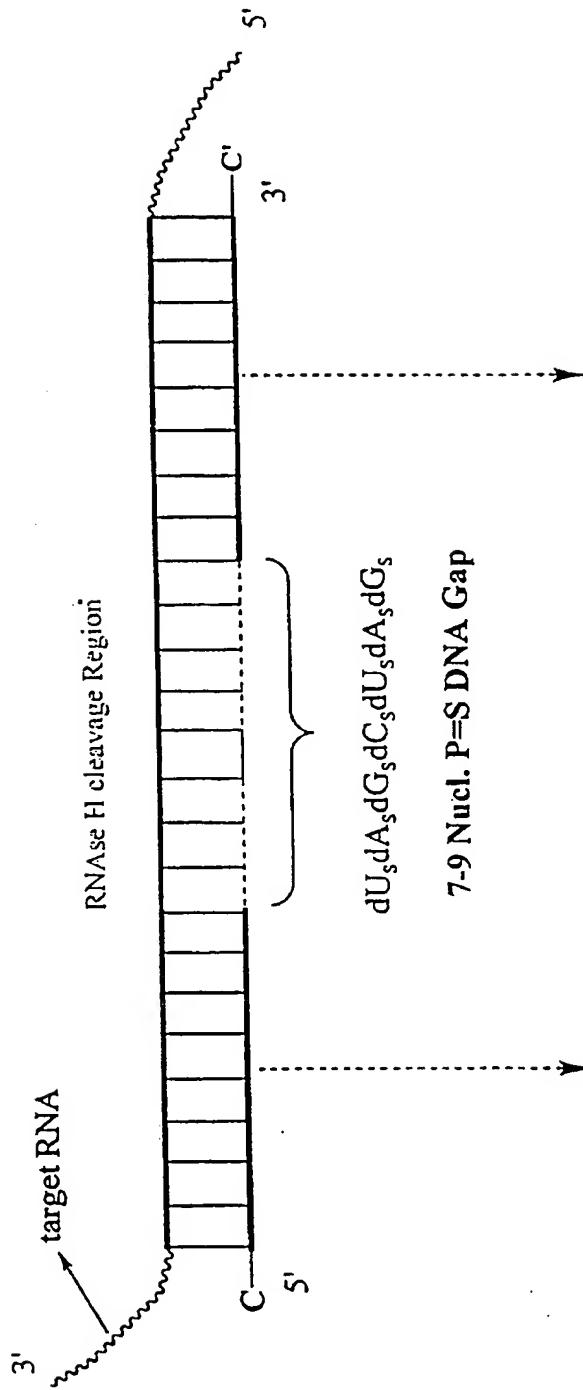
104. The nucleic acid molecule of any of claims 1-3 and 4-7, wherein (Y)<sub>n</sub> comprises a combination of phosphorothioate and 3'-S-phosphorothioate linkages.

105. The nucleic acid molecule of any of claims 1-3 and 4-7, wherein (Y)<sub>n</sub> comprises a combination of phosphorodithioate and 3'-S-phosphorothioate linkages.

106. The nucleic acid molecule of any of claims 1-3 and 4-7, wherein (Y)<sub>n</sub> comprises a combination of phosphorothioate and 3'-S-phosphorodithioate linkages.

107. The nucleic acid molecule of any of claims 1-3 and 4-7, wherein (Y)<sub>n</sub> comprises a combination of phosphorodithioate and 3'-S-phosphorodithioate linkages.

01/23



Novel nuclease resistant flanking sequences with strong hybridization properties

Fig. 1

02/23

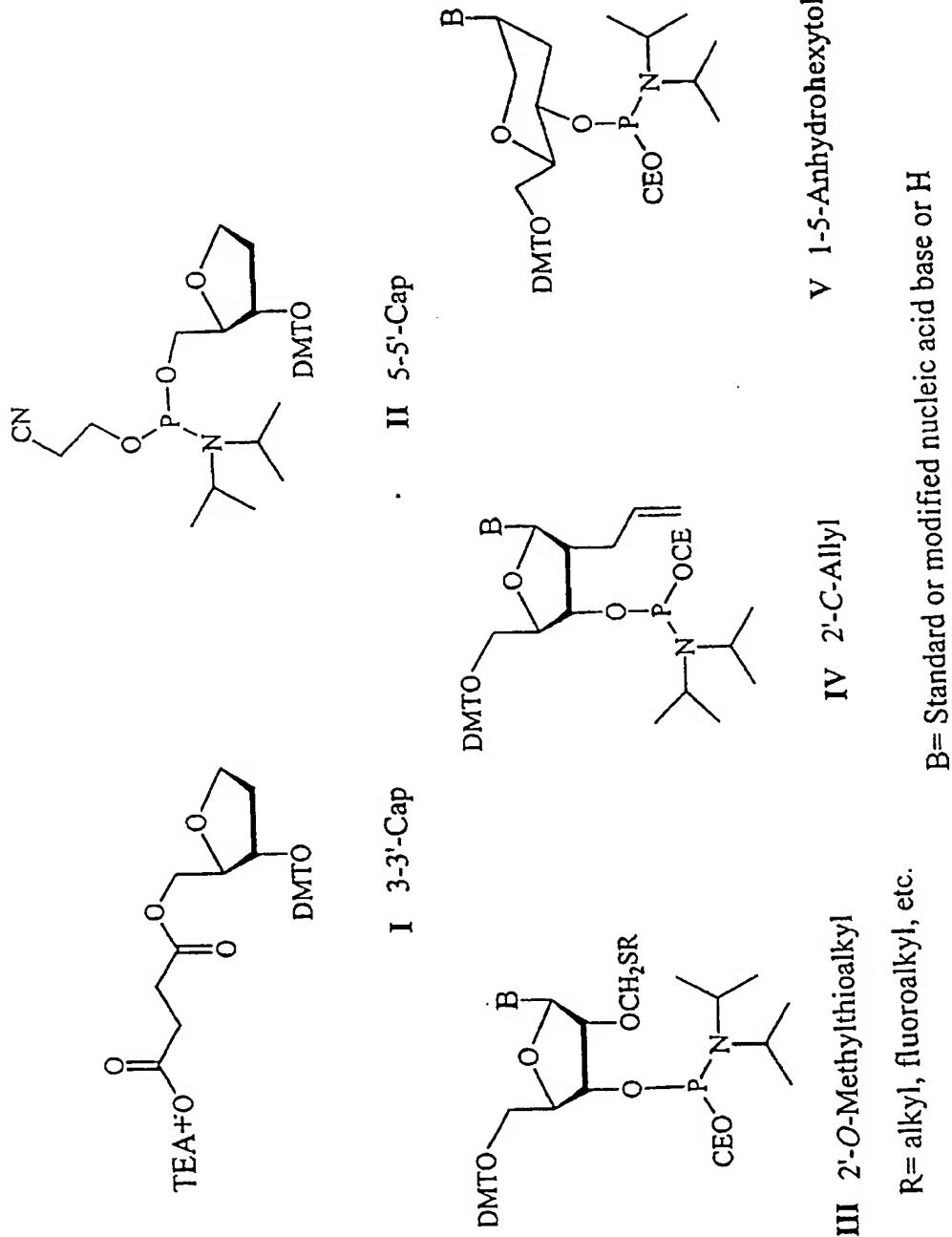


Fig. 2

03/23

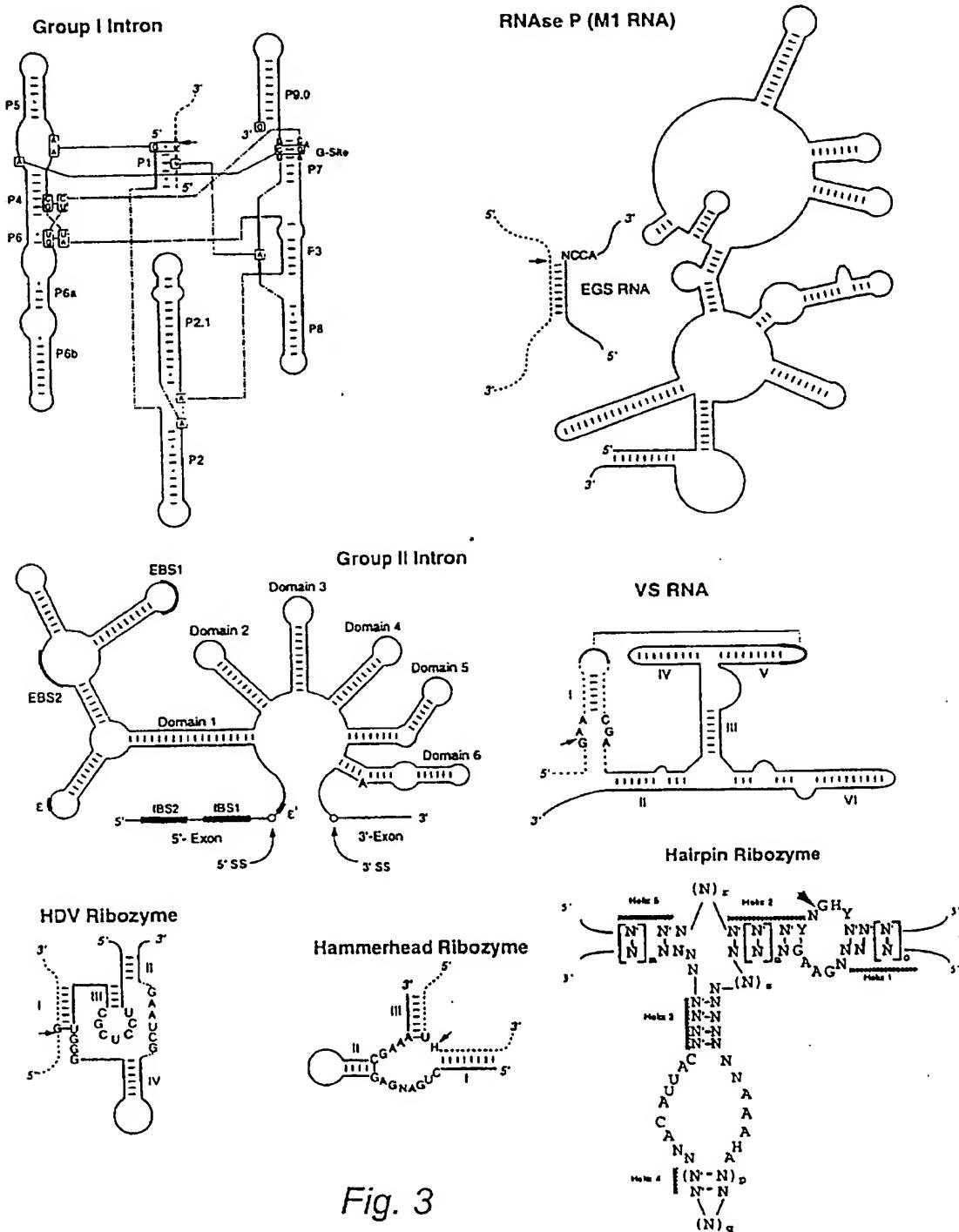


Fig. 3

04/23

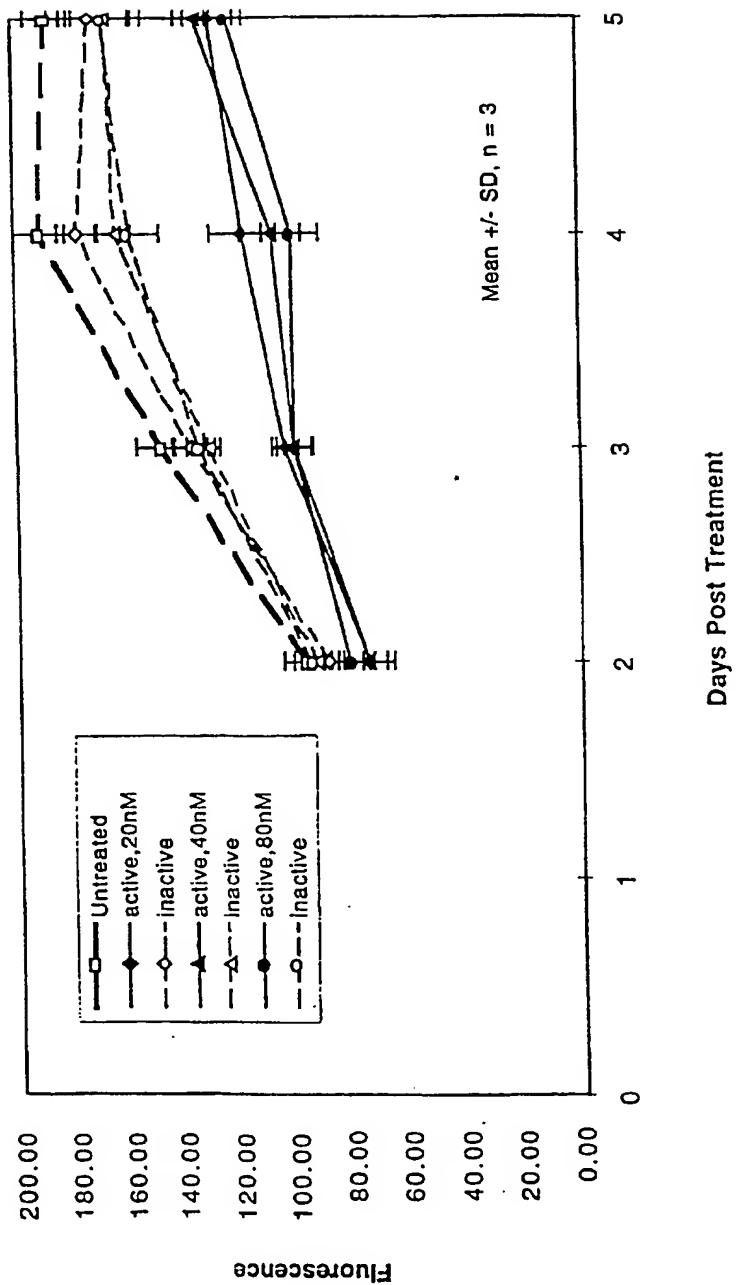


Fig. 4

05/23

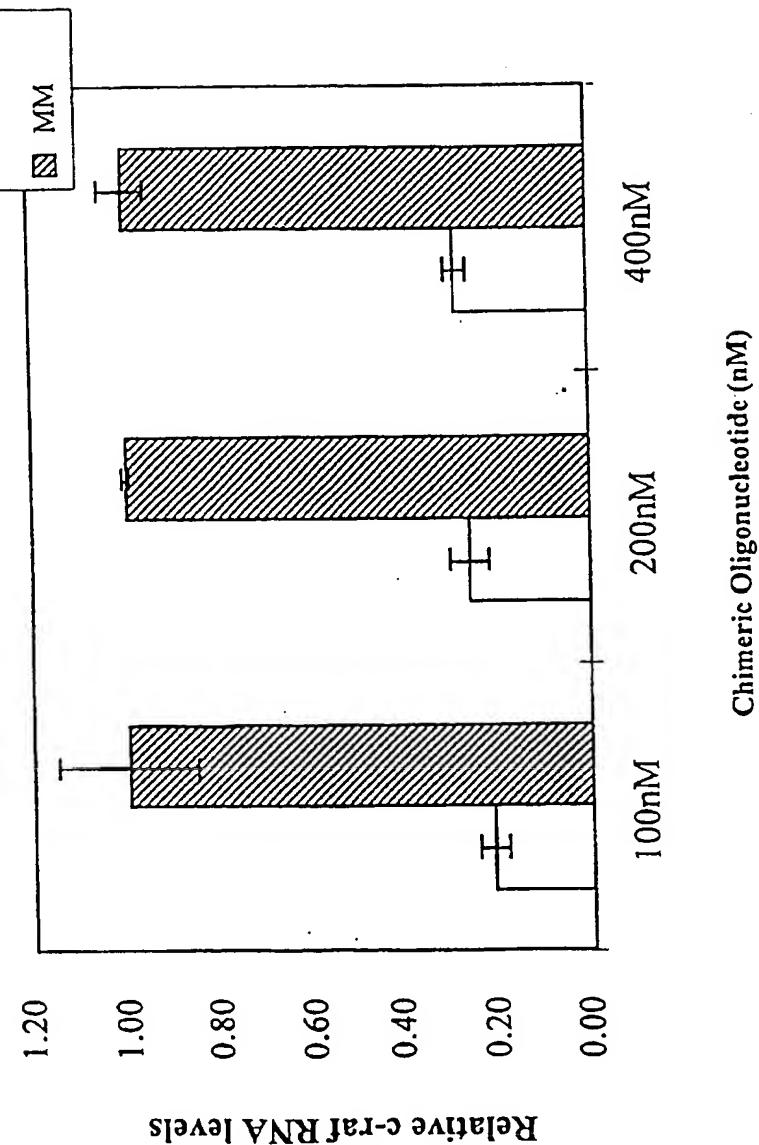
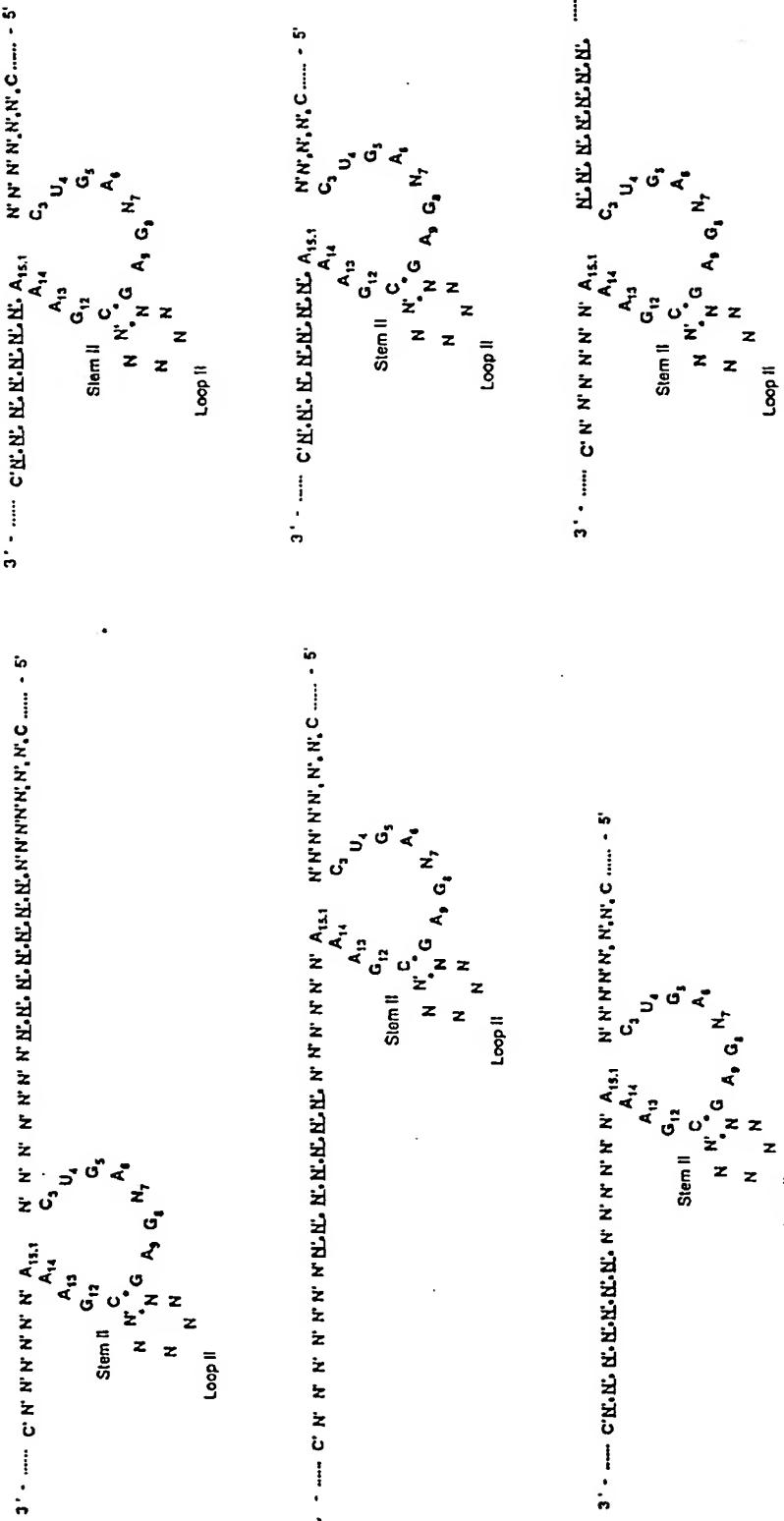


Fig. 5

06/23



07/23

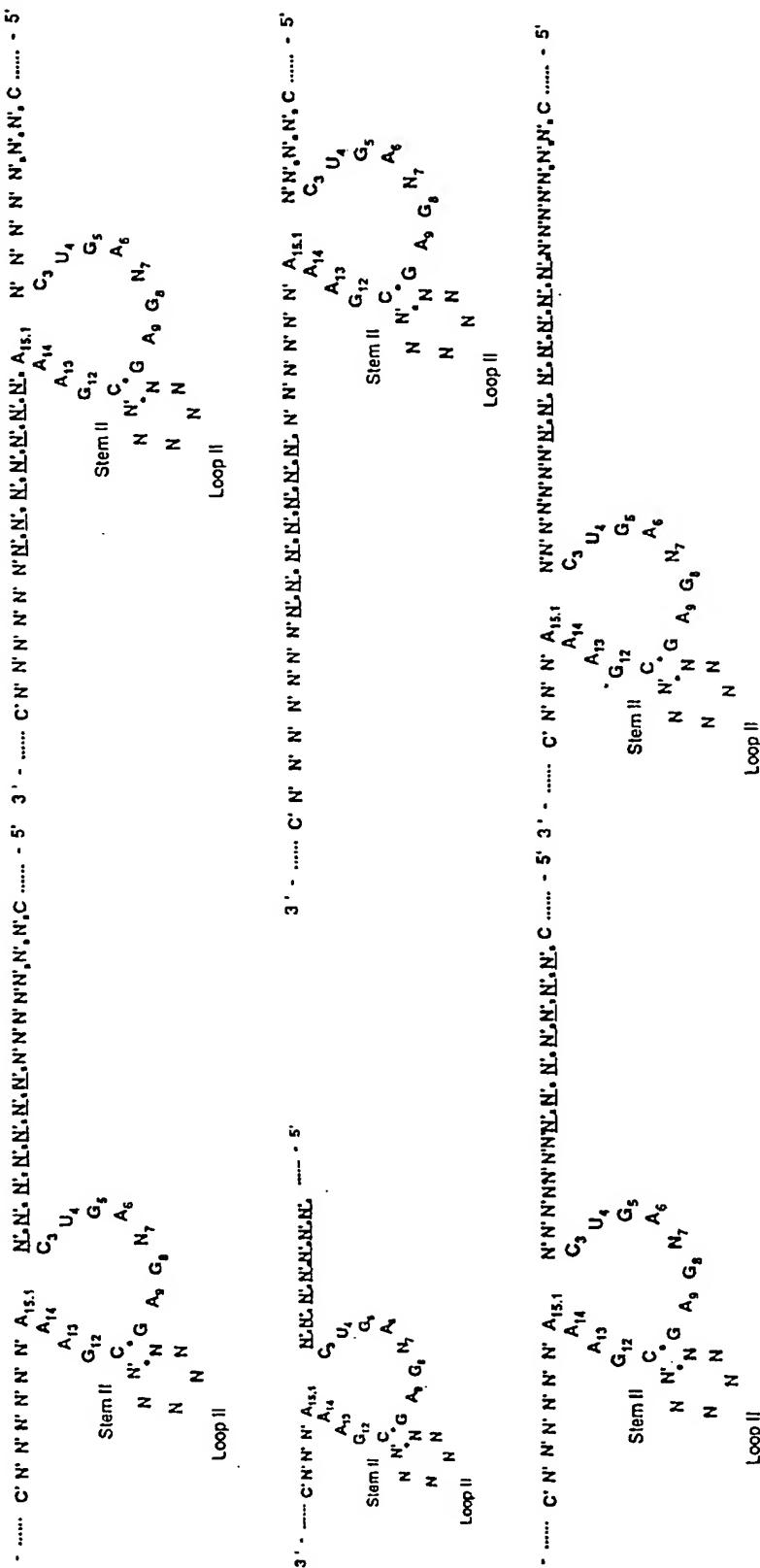


Fig. 6B

08/23

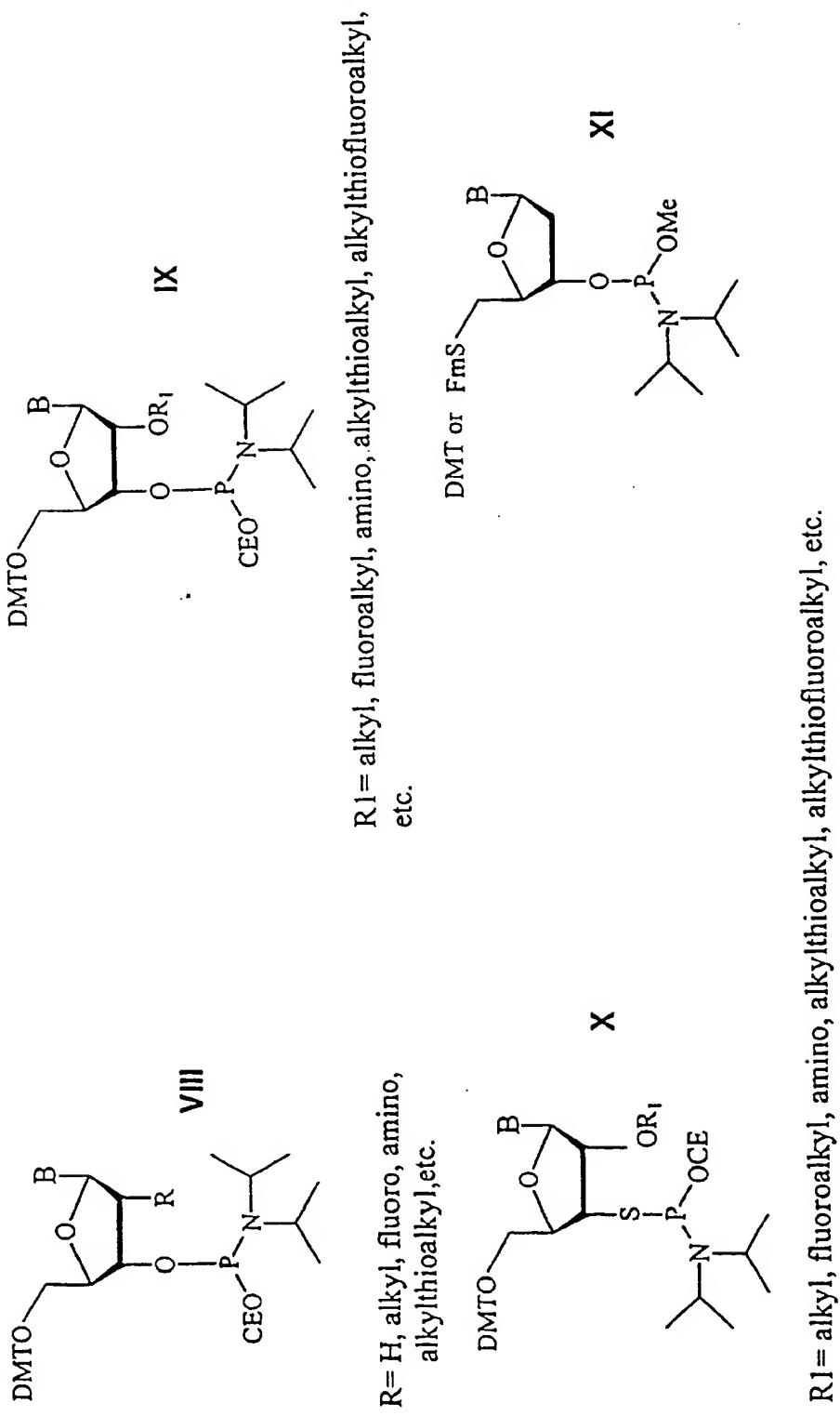
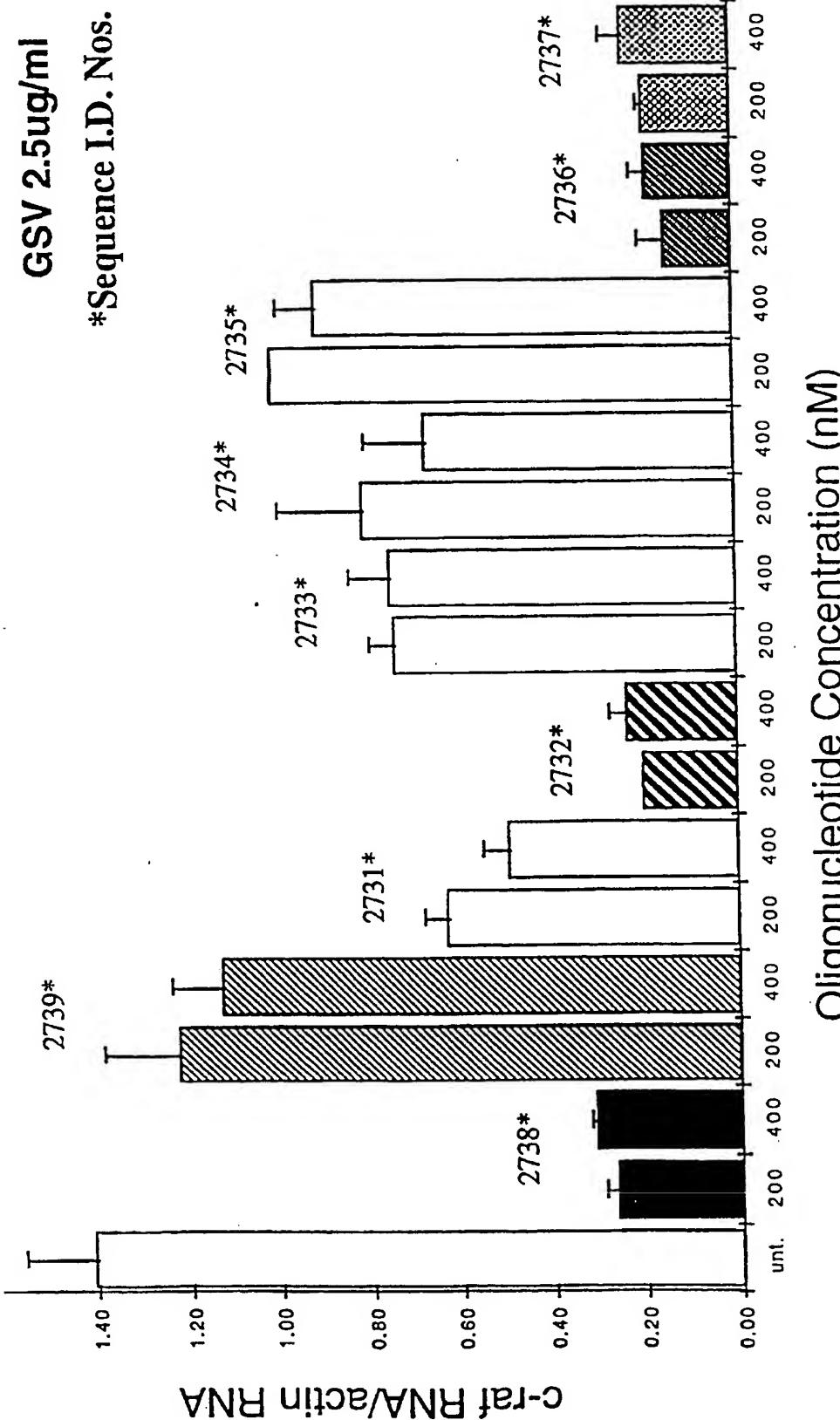


Fig. 7

09/23



SUBSTITUTE SHEET (RULE 26)

Fig. 8

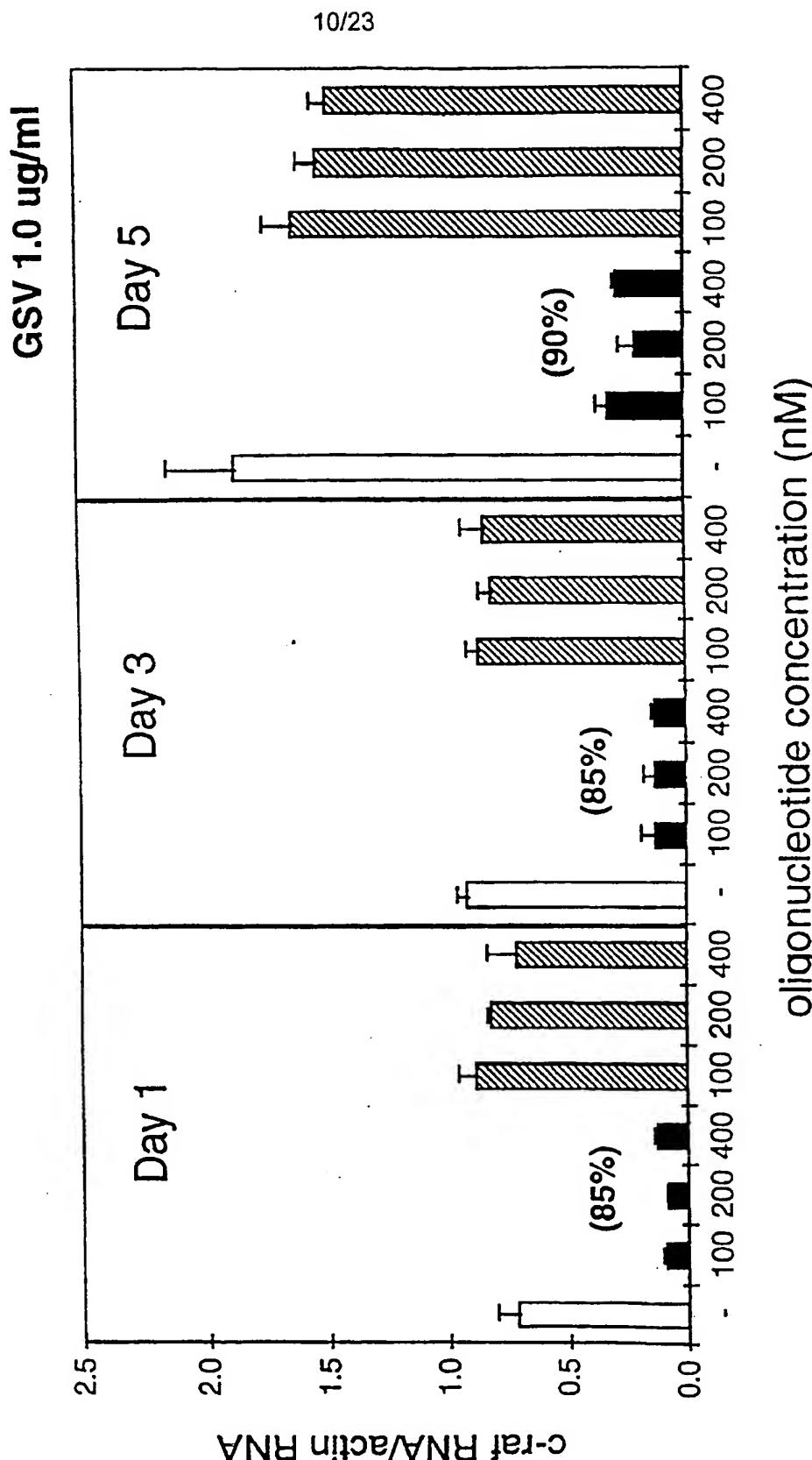


Fig. 9

11/23

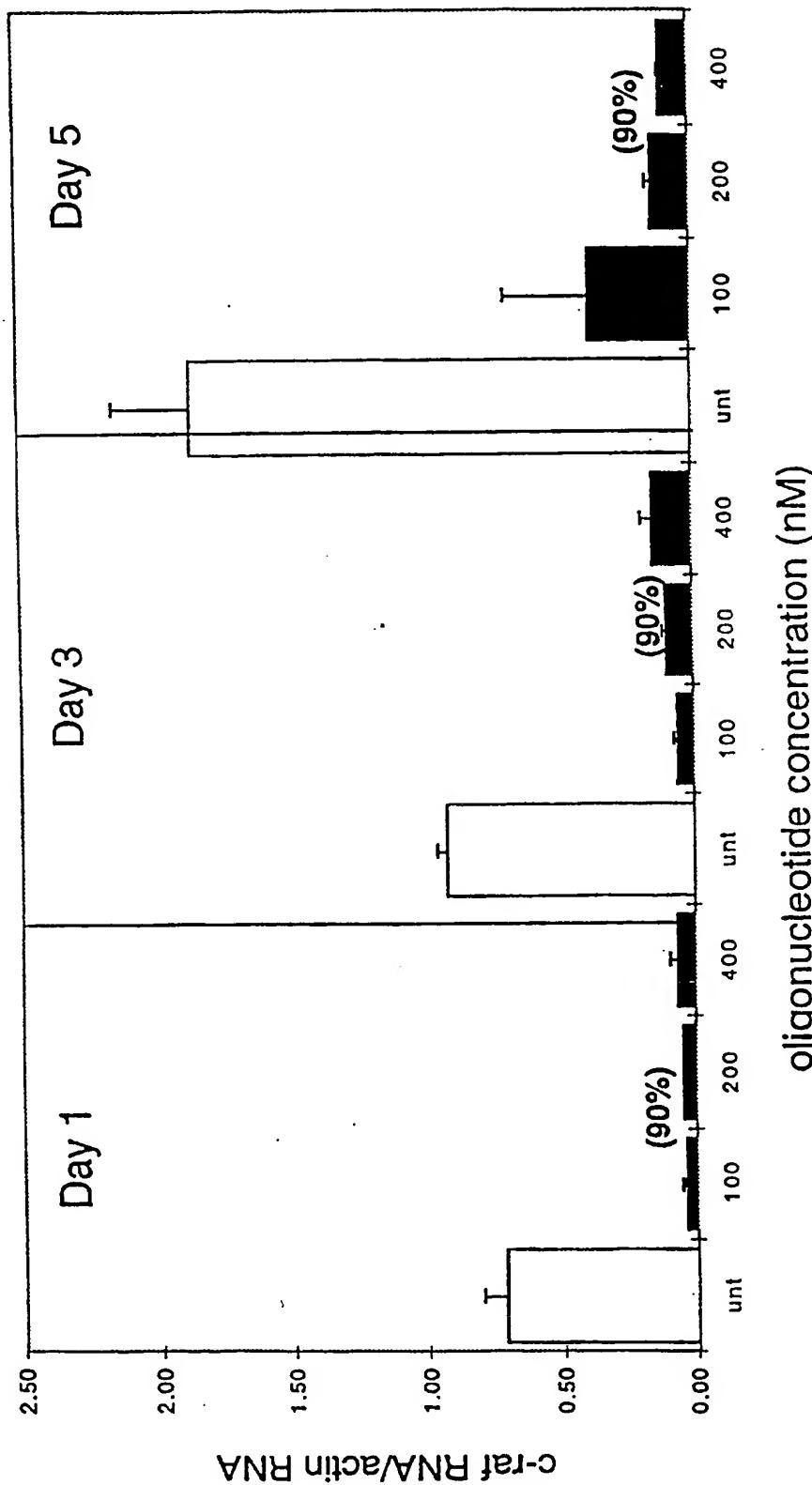


Fig. 10

12/23

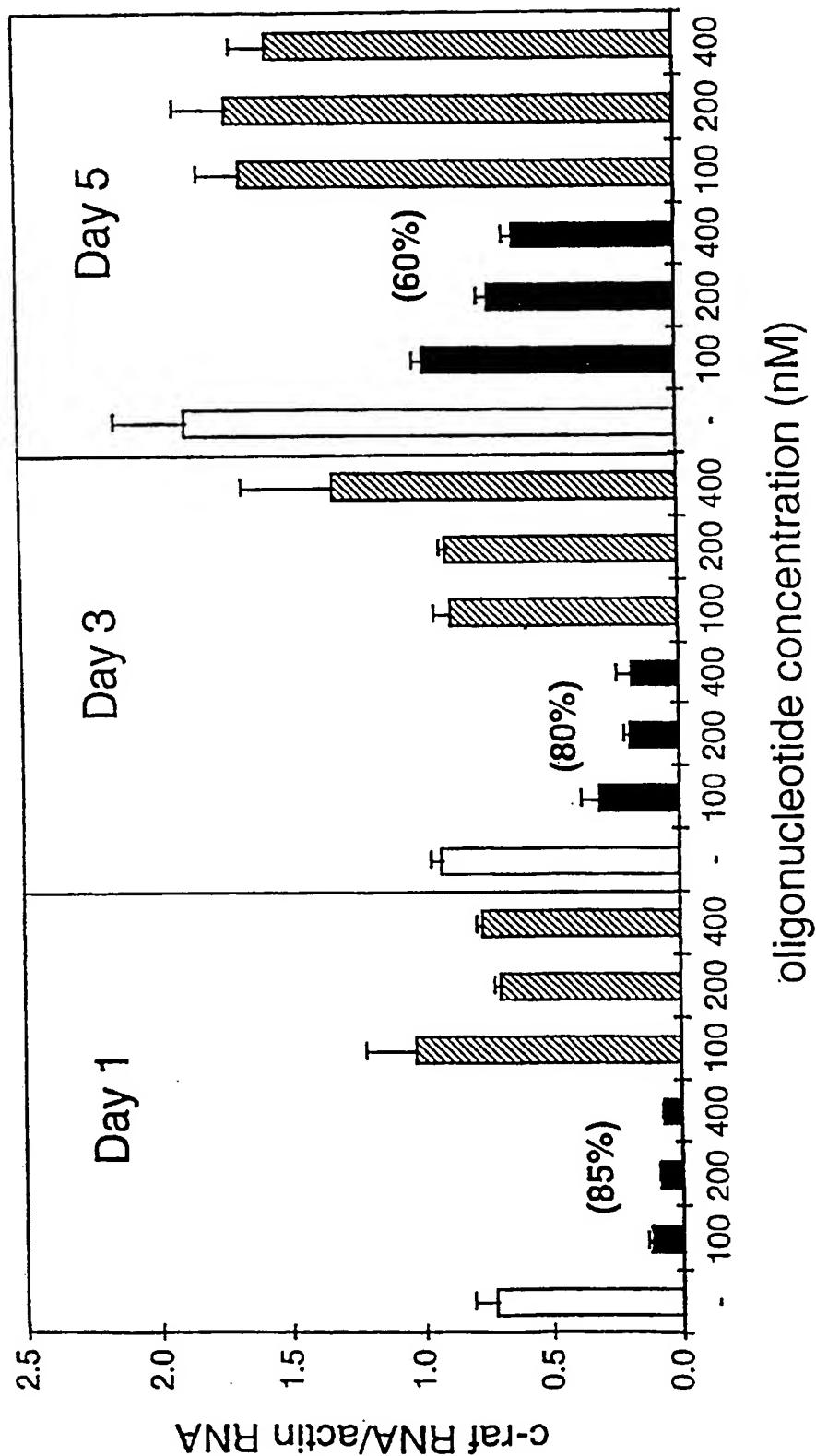


Fig. 11

13/23

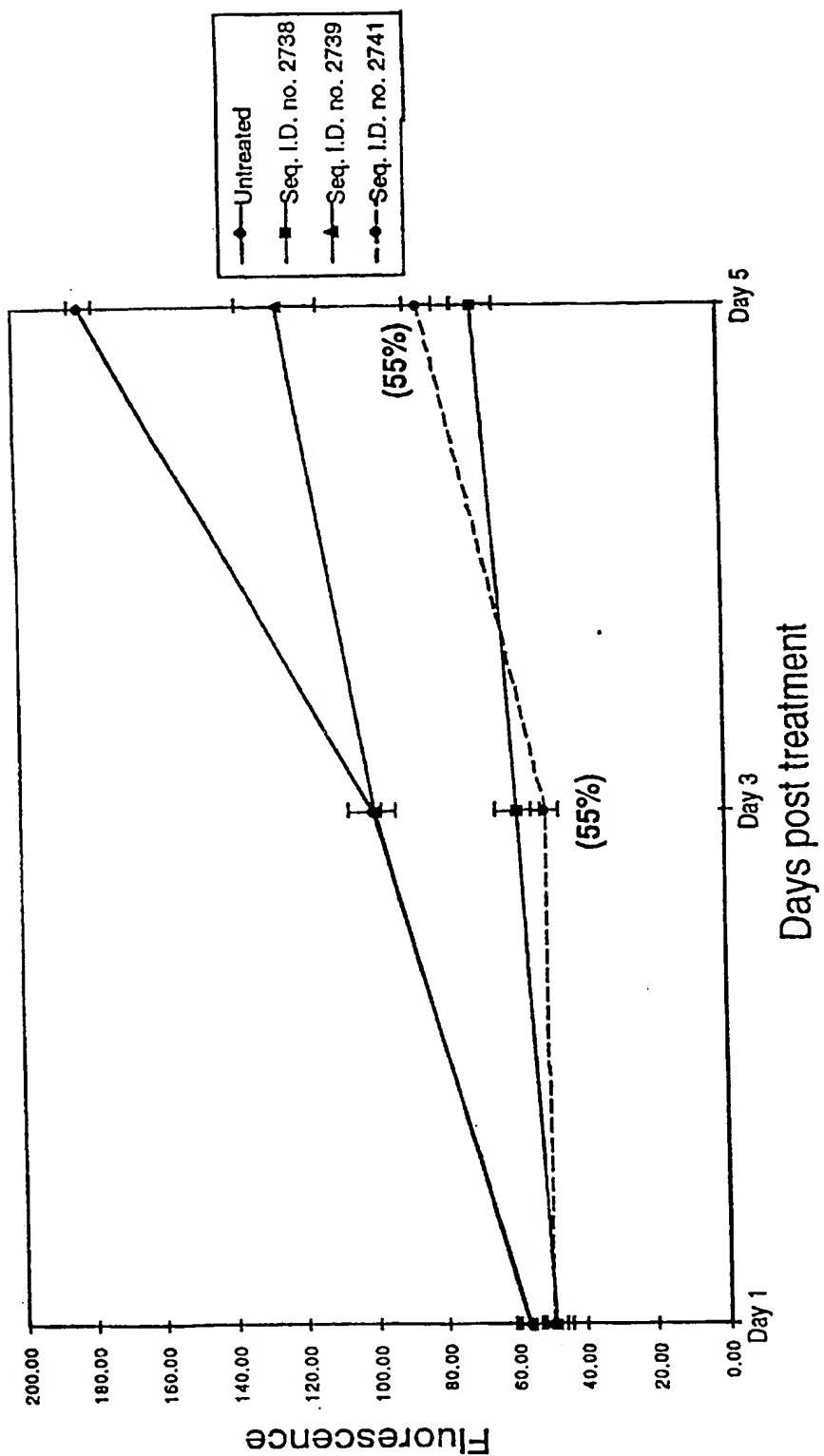


Fig. 12

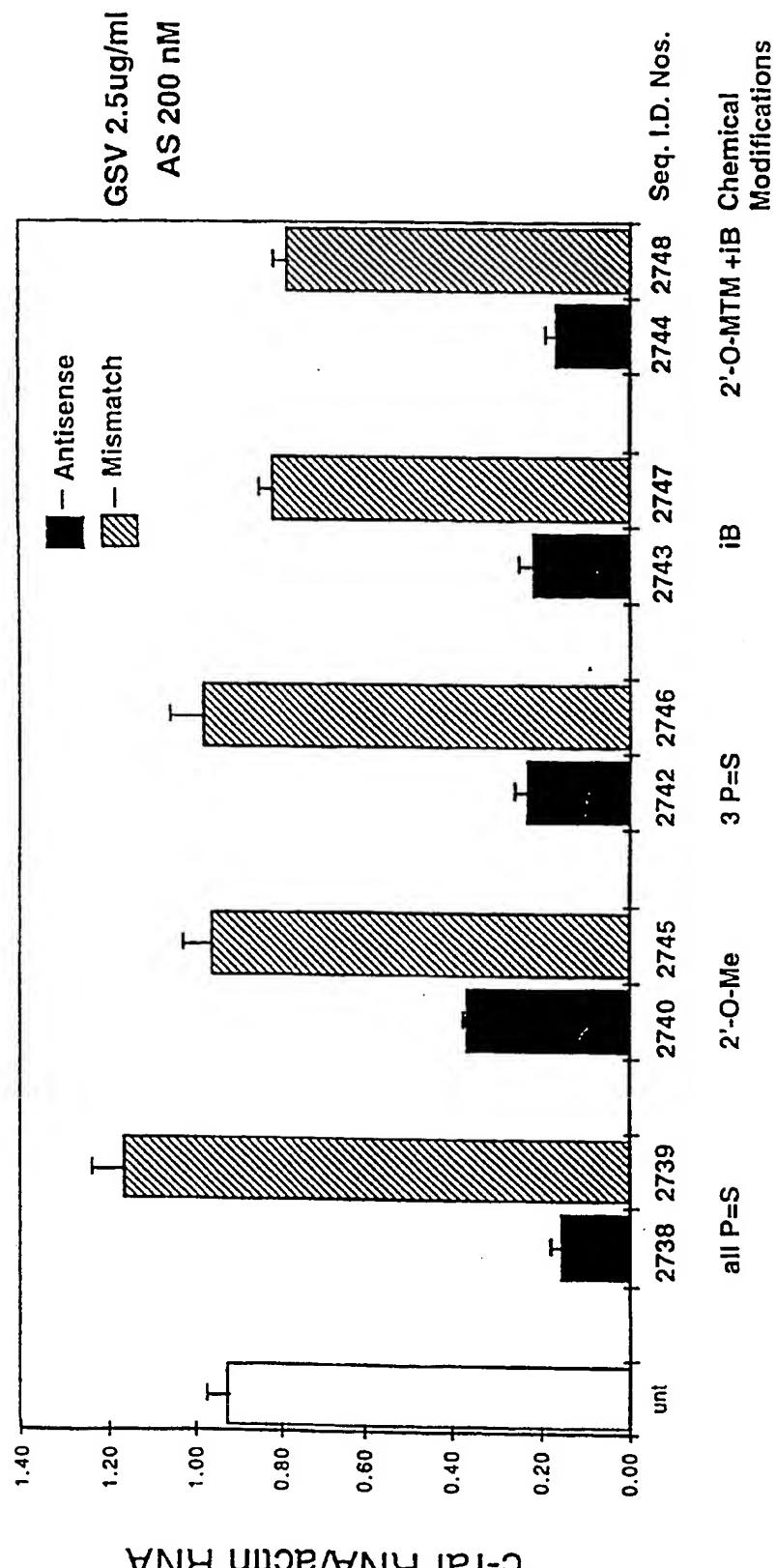


Fig. 13

15/23

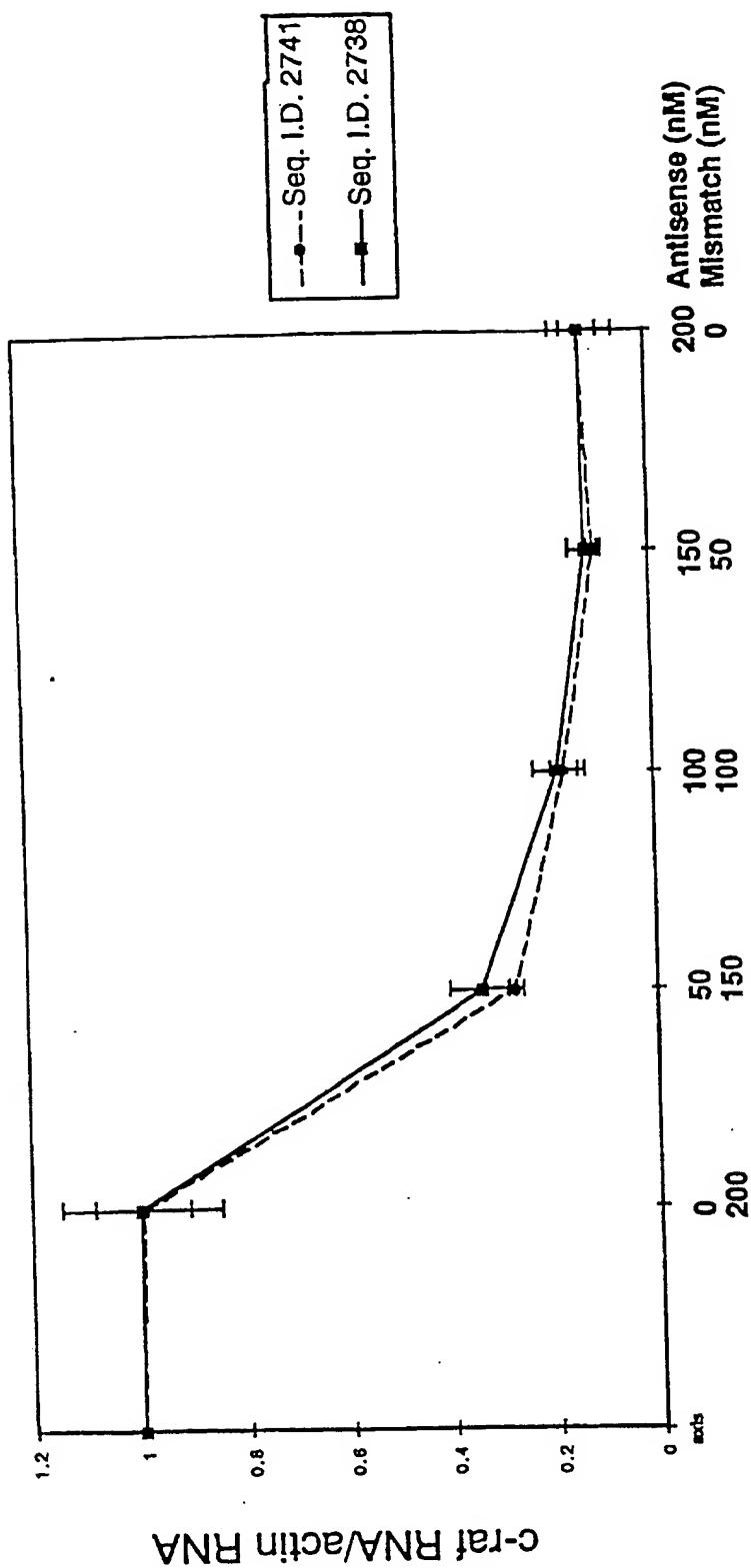
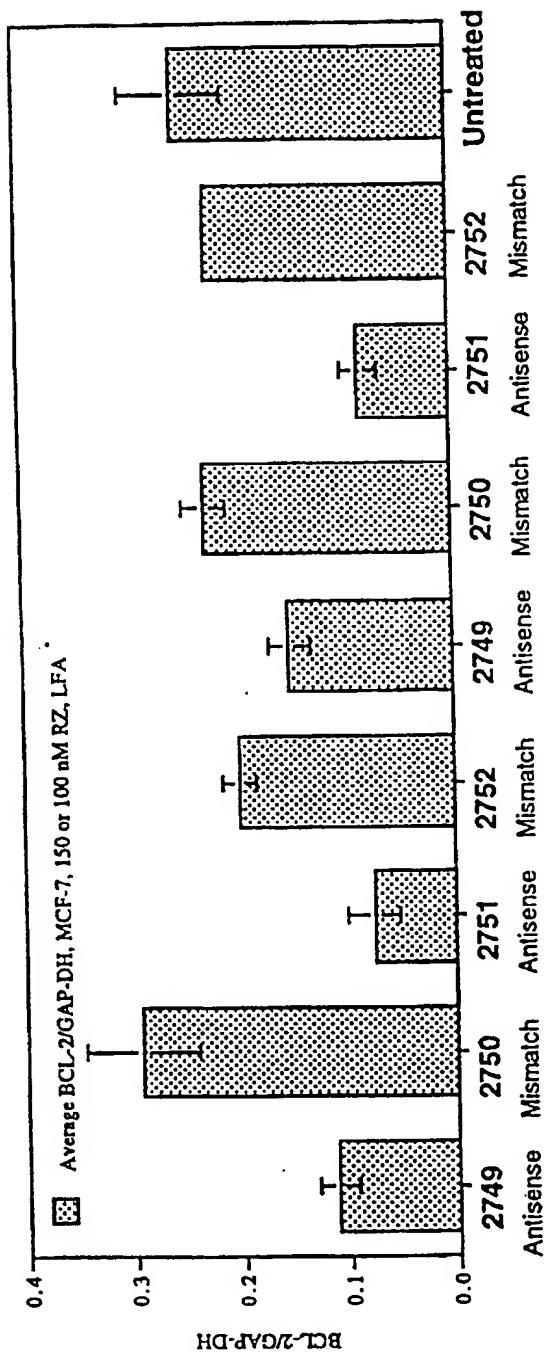
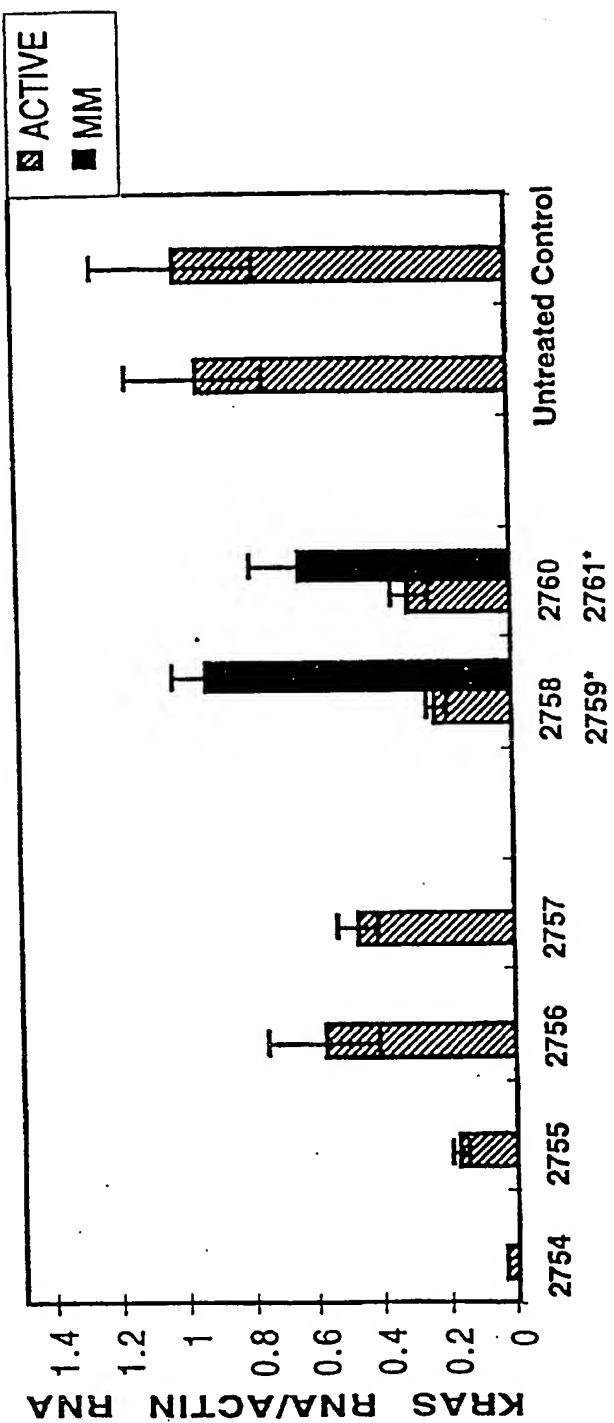


Fig. 14

16/23

**Seq. I.D. Nos.***Fig. 15*

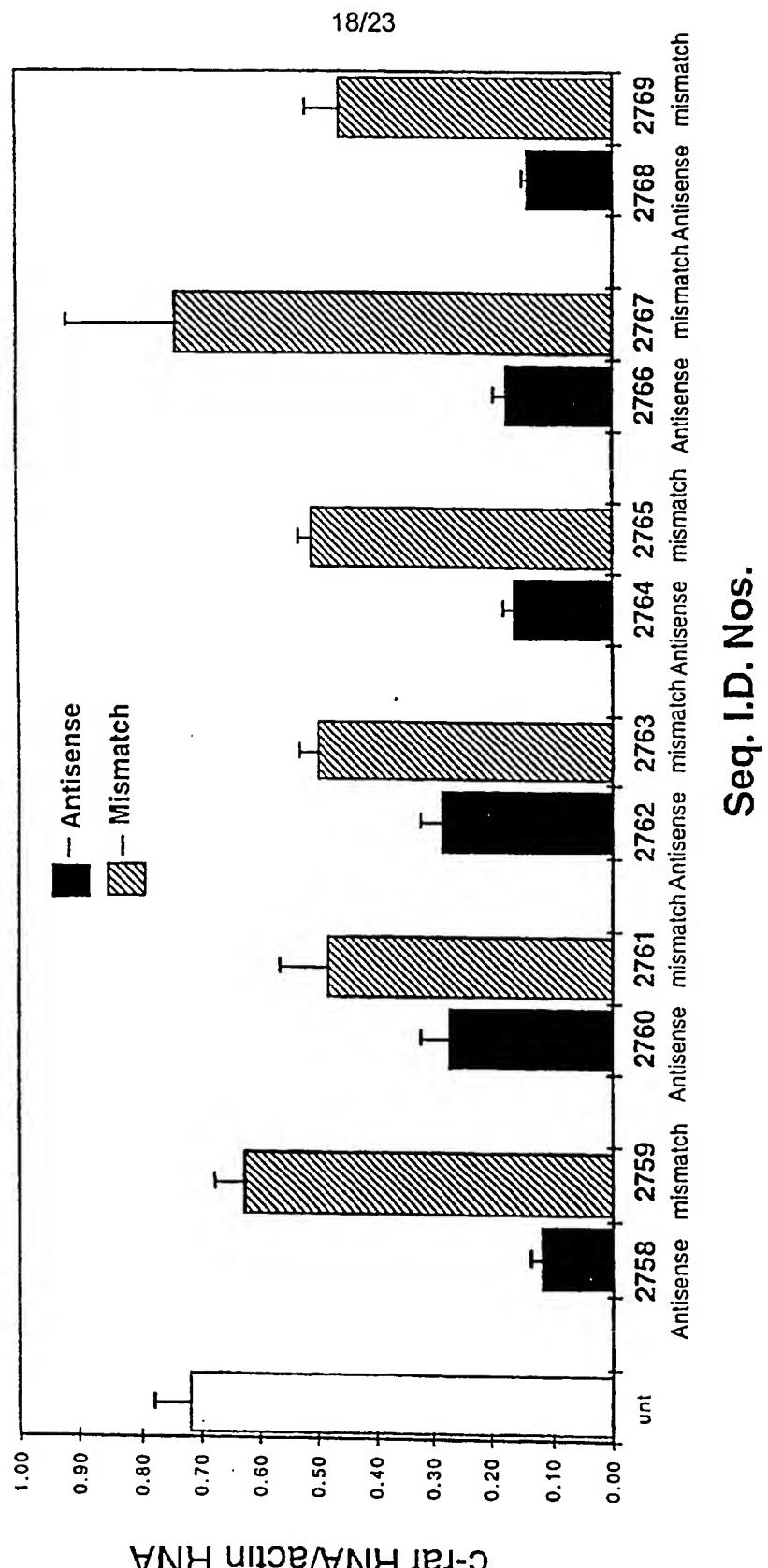
17/23



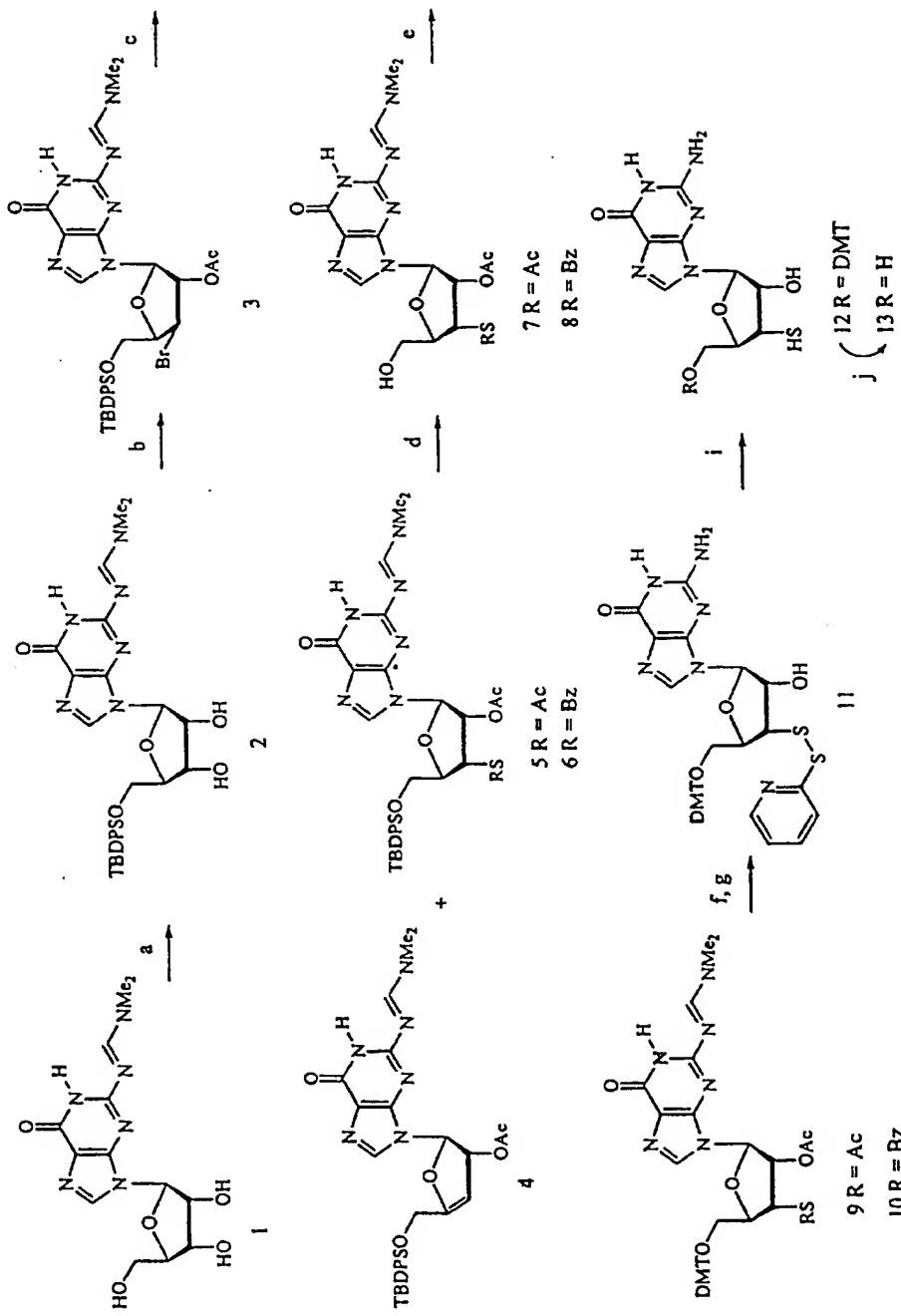
\*-Mismatch control Sequence

Sequence I.D. Nos.

Fig. 16



19/23



Reagents and conditions: (a) TBDDPS-Cl, Pyr, rt, 16 h; (b)  $\text{Me}_2\text{C}(\text{OAc})_2\text{COBr}$ , MeCN,  $\text{H}_2\text{O}$ , rt, 3 h; (c) KSAc or KSBz, DMF, 60 °C, 10 h; (d) TBAF•3H<sub>2</sub>O, AcOH, THF, rt, 5 h; (e) DMT-Cl, Pyr, rt, 4 h; (f) 40% aq. MeNH<sub>2</sub>, rt, 16 h; (g) 2,2'-dipyridyl disulfide, DMF, 60 °C, 10 h; (i) DTT, CHCl<sub>3</sub>, rt, 3 h; (j) 1N HCl in MeOH, DTT, rt, 3 h.

Fig. 18

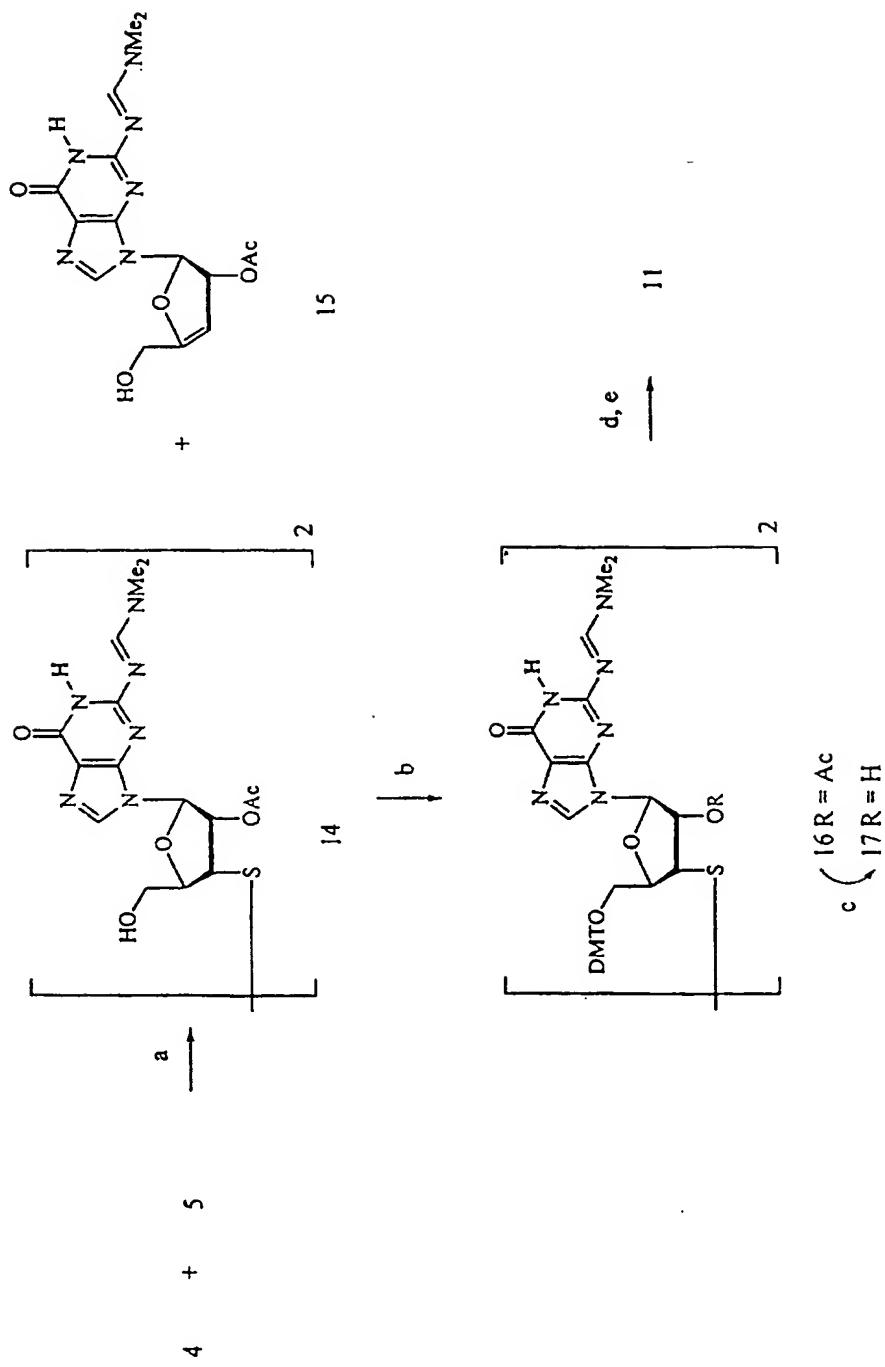
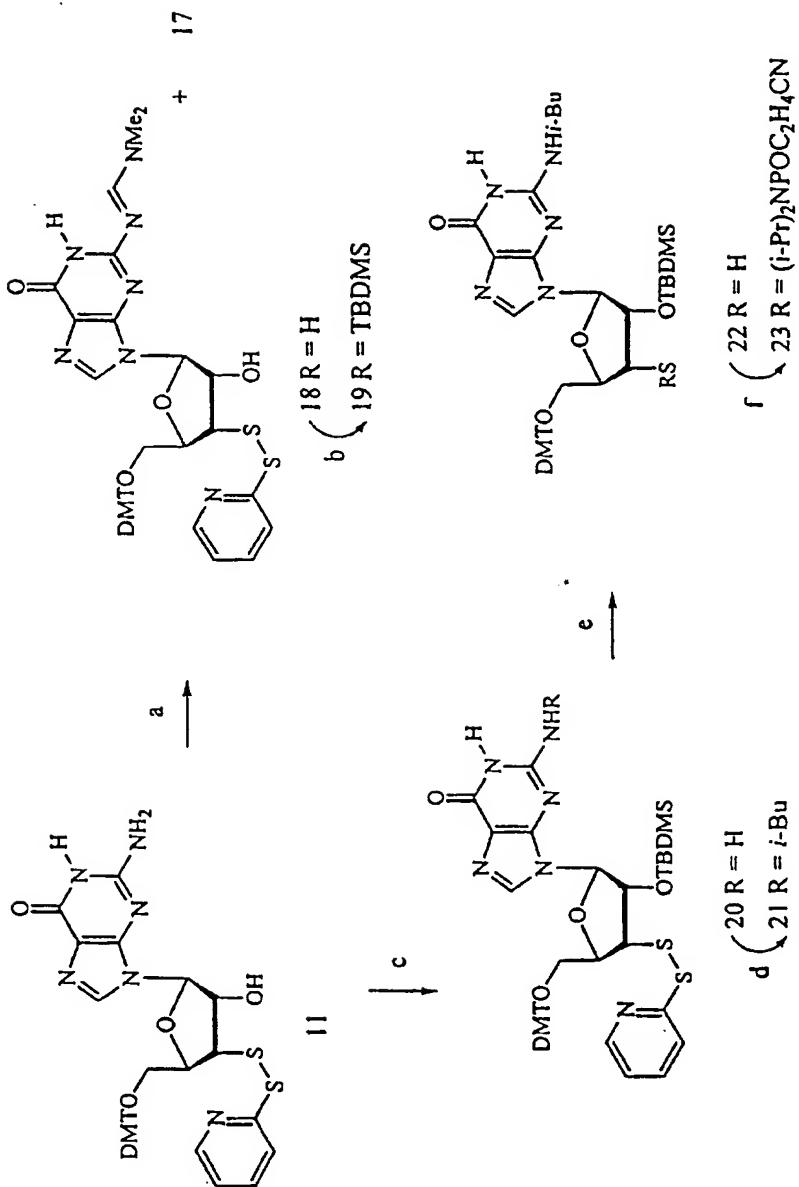


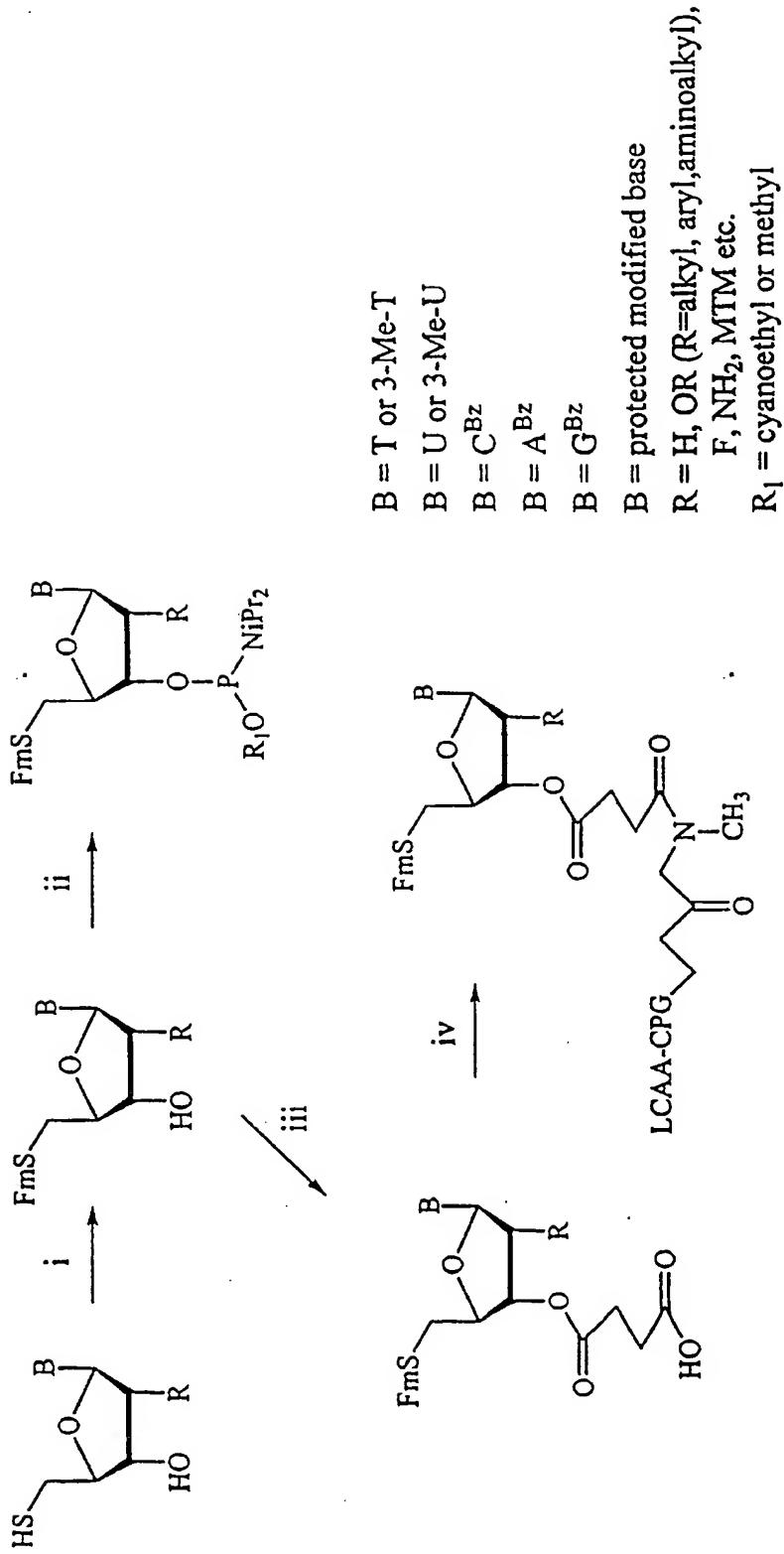
Fig. 19

21/23



Reagents and conditions: (a)  $\text{Me}_2\text{NCH}(\text{OMe})_2$ , Pyr, rt, 16 h; (b)  $\text{TBDMSCl}$ , Pyr, Im, rt, 5 h; (c)  $\text{TBDMSCl}$ , Pyr, Im, rt, 16 h; (d)  $i\text{-Bu}_2\text{O}$ , Pyr, DMAP, rt, 16 h, then 50 °C, 5 h; (e) DTT,  $\text{CHCl}_3$ , TEA; (f)  $(i\text{-Pr})_2\text{NPOC}_2\text{H}_4\text{CN}$ , DIPEA, I-MeIm, rt, 2 h.

Fig. 20



Reagents: (i) fluorenylmethyl chloride, DIEA, DMF or DCM; (ii) 2-cyanoethyltetraisopropylchlorophosphorodiimidite (for  $R_1 = \text{cyanoethyl}$ ) or methyltetraisopropylchlorophosphorodiimidite (for  $R_1 = Me$ ), tetrazole; (iii) succinic anhydride, DMF; (iv)  $CH_3NH-CH_2CONH-LCAA-CPG$ , DCC.

Fig. 21

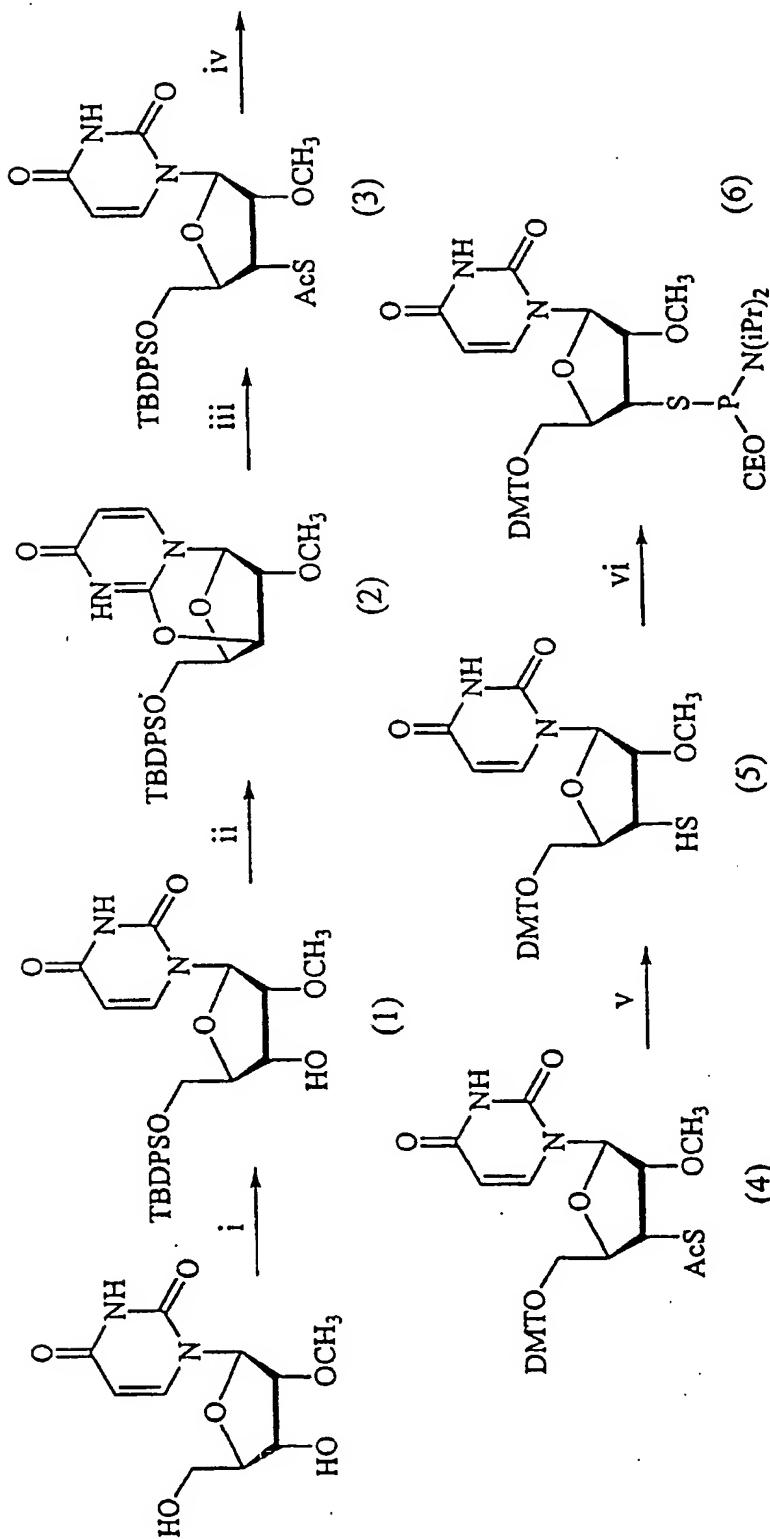


Fig. 22